JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOLUME 50, 1955

NUMBERS 269–272

Published Quarterly by the

AMERICAN STATISTICAL ASSOCIATION

WASHINGTON, D. C.

1955

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

The Editors welcome the submission of manuscripts for possible publicatio They should be typewritten entirely double-spaced, including footnotes, an two copies should be sent to the Editor, W. Allen Wallis, 207 Haskell Hal University of Chicago, Chicago 37. Books for review should be sent to the sam address. Unsolicited book reviews are not accepted, but suggestions of titles fc review are welcome.

Corrigenda: Readers and authors are urged to submit to the Editor notices o errors found in this or any previous volume. These will be published onc a year, in the December issue.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

CONTENTS OF VOLUME 50

INDEX TO VOLUME 50, 1955

ARTICLES

BOOK REVIEWS

LIST OF REVIEWERS

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 269 MARCH 1955 *Volume 50*

CANADA—NORTHERN NEIGHBOR*

HERBERT MARSHALL
Dominion Bureau of Statistics, Canada

The subject of my address is "Canada, Northern Neighbor." Since, recently, there has been an intensification of interest in Canada on the part of your countrymen, I came to the conclusion that the subject I have chosen might have a more general appeal than any other I might select.

Canadians in recent years have been exercising their critical faculty on the broader aspects of national identity and characteristics. Some have wondered if we really have developed a national consciousness, others have described us as "living under the tyranny of the Sunday Suit," as conservative, self-conscious, lacking self-confidence One, perhaps with tongue in cheek, for he has been termed Canada's literary gadfly, declares we possess traits of infantilism in our national life. We took heart, however, from a later critic who advanced us to the stage of adolescence. If Canada has reached her present status while still no more than adolescent, surely we may have hopes for a satisfactory maturity!

Of late there have been great happenings in Canada; spectacular discoveries in natural resources and industrial developments with important implications for both the United States and Canada which could not fail to draw attention to the affairs of the Northern Neighbor.

Canada possesses more square miles of territory than any other nation save the Union of Socialist Soviet Republics and China. Its area is 3,846,000 square miles but a mere statement of superficial area can be misleading. The vast open spaces of Canada have not seldom attracted the envious gaze of nations whose population requires more space. In Canada as a whole there are fewer than four persons per square mile of territory which compares with over fifty in the U.S.A. The Yukon and Northwest Territories, not yet organized into prov-

* Presidential Address at the 114th Annual Meeting of the American Statistical Association, Montreal, Canada, September 11, 1954.

inces, which comprise over 1½ million square miles or 39% of the surface of Canada, had only 25,000 population in 1951. It is the nature of these sparsely settled areas and the northern parts of the Canadian provinces which accounts for the low population density of the Dominion as a whole.

An outstanding physical feature of this country is the Canadian Shield which has had and will continue to have a profound influence on our development. It is a huge U-shaped area of Precambrian rock, interspersed with sedimentary and volcanic intrusions, comprising 1,800,000 square miles or 45% of our total territory. It surrounds the great inland sea known as Hudson Bay and extends from the Coast of Labrador on the Atlantic, to the Interior Plains of the Prairie Provinces, and at one point stretches down to the Thousand Islands in the St. Lawrence River At its top it is 1,900 miles across. It covers most of Quebec, a very large though somewhat smaller portion of Ontario, three-fifths of Manitoba, one-third of Saskatchewan, the northeast corner of Alberta, approximately one-half of the Northwest Territories and thrusts up in places into the Arctic Islands.

The Canadian Shield is not suited for agriculture except in scattered pockets If it had been our population would be much larger than it is. It is a treasure trove of mineral wealth; its rushing rivers are a source of electric power; it has vast forested areas, the source of pulp and paper and lumbering industries, and contains many fur-bearing animals, the natural beauty of its scenery, its lakes and rivers, many of which abound in fish, are a tourist attraction with great drawing power.

If the Canadian Shield is 45% of our territory, what of the other 55%?

Time does not permit of a detailed description of the other regions. In broad outline it may be said that the south-eastern part of the province of Quebec, the three Maritime provinces, Nova Scotia, New Brunswick, and Prince Edward Island, and Newfoundland are in the Appalachian region. Therefore that area is cut up by mountain ranges which limit the amount of arable land There is no coastal plain but there is the submerged grand banks, one of the great fishing areas of the world

In the central provinces, Quebec and Ontario, the region of heavy settlement is narrow because the Canadian Shield reaches far to the south. The total area of these two provinces is over one million square miles. A section to the south along the lower Great Lakes and the Upper St Lawrence, including the Niagara Peninsula, comprising about 50,000 square miles, is known as the St Lawrence Lowlands It contains between one-half and two-thirds of the population of Canada and is the industrial heart of the Dominion

From the western limit of this Section, that is, from Lake Huron

westward, there is a stretch of a thousand miles before the continental plain is reached. That thousand miles is occupied by the southward thrust of the Canadian Shield and has been a great barrier to east-west communication Even to-day Canadians motoring, say, from Montreal to Vancouver, avoid this stretch by going in transit through the United States. Work is progressing on a Trans-Canada highway. In the course of a rail journey through this region the passenger will look out upon hundreds of miles of wilderness with only occasional centres of settlement.

The Continental Plain occupies the greatest portion of the prairie provinces, that is, the area not taken up by the Canadian Shield. It has deep arable soil from which the bulk of Canada's grain is produced. But these provinces, especially Alberta, have come into the limelight recently because of oil and gas discoveries.

In the extreme West is the Cordilleran Region. Here high mountain ranges are interspersed with valleys where mixed agriculture and fruit growing are important Other basic industries are mining, lumbering, and fishing

North of the Prairie Provinces and stretching to the Arctic Ocean lie the Yukon and Northwest Territories. This huge area presents a variety of physical conditions; treeless plains in the far north; the rolling hills of the Canadian Shield; the forested valley of the Mackenzie river. Water resources vary from small streams and lakes to the largest rivers in Canada, the Mackenzie and Yukon, both over 2,000 miles in length, and Great Slave and Great Bear Lakes, each of which is 11,000 square miles.

Canada obviously is a land with a difficult topography. There is no possibility of continuous settlement throughout all its regions. At present Canadians are concentrated in a line some 4,000 miles long with thinly populated sections intervening and the great bulk live within 200 or 300 miles of the border.

Canada also has difficult climatical conditions characterized by hot summers and cold winters with heavy snowfall The climate of southern British Columbia is an exception resembling to some extent that of England. Northern Canada, of course, has Arctic conditions and no sections of Canada have a climate corresponding to that of southern Europe.

We have survived and developed as a nation, not only in spite of topography, geography and climate, but in spite of other influences which might have caused another story to be written. In 1760, when Wolfe won his victory on the Plains of Abraham, this upper part of North America passed into the hands of Britain. The policy adopted towards the French inhabitants shortly after was important for our

future. They were granted full freedom of worship, speech, and customs. This was probably an important reason why they did not favor joining the newly formed United States after 1776 despite the fact that the colonies of British North America were small, isolated and relatively weak. This decision to remain in the British orbit created a refuge for those in the new Republic who had opposed the Revolution, and we had an influx of what have been called United Empire Loyalists who have contributed much to the development of Canada. Thus, in the very early stage of our history, our big neighbor has strongly influenced our development. Naturally the United States has continued to influence us in many ways Most of this influence has been good and benign but it has not been invariably so. There were the border skirmishes of 1812 and raids at later dates There was the threat of your westward surge and the doctrine of Manifest Destiny. But these events spurred Canadians to accomplish what to some seemed impossible in order to preserve the nation. The first was the achievement of Confederation by which, in the face of much opposition, four provinces entered the federation. These were Nova Scotia, New Brunswick, Quebec and Ontario, the two last having been Upper and Lower Canada previously. This was in 1867, and in 1872 Prince Edward Island joined. Away on the Pacific Coast was another colony—British Columbia. Between the settled parts of Ontario and British Columbia stretched a practically empty country for two thousand miles, including what was to become the great grain growing provinces of the Prairies. British Columbia's condition for entering Confederation was the building of a railway across these empty spaces

Through miles of Precambrian rock, through regions that seemed to be more water than land, across the prairies and through the three ranges of the Rocky Mountains the railroad was pushed and the first train reached Vancouver in 1886 Thus Canada became bound together by a thin line of steel. It was many years afterwards that the great open spaces of the Prairies really began to fill up with people instead of herds of buffalo

The survival of Canada as a nation, in spite of its handicaps in topography, geography, and climate and in spite of the siren allurements of your prosperity, your big markets, and sometimes your pressures, is one reason why historians declaim on the "Miracle of Canada." There are others To create a separate nation we had to force our economy to a large extent into an east-west development, whereas, according to nature in some sections at least, it should be north-south. This is particularly true of the extremely west and east provinces.

Canadian destiny, of course, has been influenced by two great powers, Great Britain as well as the United States. Canada, like the United

States, has a federal organization but, apart from the federal aspect, Canadian constitutional practice is modelled closely on the British Parliamentary system.

Ours is the Parliamentary Cabinet System of government. The Cabinet, headed by the Prime Minister, brings in most of the legislation and can, except in very unusual circumstances, count on the support of the members of the party. We have a Senate in which members are appointed by the Government for life. It is very different from the U.S. Senate. In fact, it plays a minor role in the affairs of the nation.

These and other basic institutions including our judicial system we have inherited from the United Kingdom. It is our belief that they give soundness and stability to our country, which make it attractive and safe for other countries to invest in. There are no obstacles to bringing capital in or taking it out.

Canada, a member of the British Commonwealth of Nations, has equality of status with the United Kingdom in all her domestic and foreign affairs. Canada is a sovereign autonomous state Some of my acquaintances in the United States have been surprised and almost skeptical when I told them we pay no taxes to the United Kingdom. Our full-hearted allegiance to England's Queen as Queen of Canada carries with it no recognition of England's government as our government. The concept of a Queen of the Commonwealth seems to many to connote some financial obligation. Doubtless the fact that Queen Elizabeth is Queen of Canada, and Queen of each Commonwealth Nation, as well as Queen of the United Kingdom, is something of a puzzle to those not directly concerned with the evolution of this political achievement. It has even been suggested that the Queen might live part of the time in the Capital of each of the Commonwealth nations.

In this connection, I quote from Professor Lower's book "Canada: Nation and Neighbor"—

> The British Empire of the last century, which was both empire of domination and empire of settlement, began to change its nature as Canada grew to maturity. The history of Canada's relation with Great Britain is not only the history of self-government of a colony of settlement but it is also the history of the reshaping of imperial institutions to reconcile them with self-government. Canada was determined to have the fullest measure of self-government; she was also determined to have it without disrupting her association with the mother-country. The result was the change, so frequently described "from Empire to Commonwealth." Canada may not unfairly claim to have been the principal architect of the British Commonwealth.

At the end of nearly a century of existence as a nation Canadians have some justification for pride in their achievements. The accomplishments of the past and the scope and magnitude of post-war develop-

ments have aroused high hopes for the future. Canadians believe that Canada has a rendezvous with destiny.

From 1946 to 1953 Canada has had an outstanding period of development. Our net gain in population during the period was more than 2½ million New investments in fixed capital were over 30 billion dollars. From 1946 to 1953 twenty-one per cent of our Gross National Product went into capital investment.

Though capital expansion commenced immediately after World War II for reequipment, reconversion and modernization, needs for which had accumulated during the war economy and to some extent in the thirties, it soon went beyond the stage of catching up. New developments were bursting out in all directions There were great projects for the utilization of resources including power, minerals, and wood products, the discovery of major oil fields in the West, the mechanization of agriculture which amounted to a revolution in farming methods and expansion in various lines of manufacturing.

Canada's development as an urban and industrial economy is evident in the fact that in 1900, 40 per cent of her labour force was engaged in agriculture In 1946 the proportion was 25%, in 1953 it was 16%

Along with this high rate of economic activity and expanded productive capacity, there has gone a considerable rise in the standard of living The production of goods and services—the gross national product—has increased after correction for price changes by one-third from 1946 to 1953 In the United States the increase was 29% Canadians enjoy a material standard of living second only to that of the United States

An outstanding aspect of our recent progress is the development of natural resources It is almost certain that we shall eventually be self-supporting in oil on balance Pipe lines have been constructed which take the oil to the Pacific Coast and to Eastern Canada markets. At the same time tremendous supplies of natural gas have been developed. A 2200-mile all Canadian pipe line is to be constructed from Alberta to deliver natural gas to Ontario and Quebec This, it is claimed, will be the greatest installation of its kind in the world. Production of natural gas has risen from 48 billion cubic feet in 1946 to 101 billion in 1953 Disposable reserves are estimated to be over 15 trillion cubic feet.

These developments are altering profoundly the prairie economies. The three prairie provinces, situated in the west half of Canada, have been predominantly agricultural through not entirely so. Gas and oil

are bringing diversification not only as primary materials but in Alberta have given rise to an important development in petro-chemical industries.

What is happening in our prairie provinces is only one example of the important developments which have been taking place across Canada and in our far northern areas They have been the subject of many articles in American periodicals some of which I am sure you have seen.

The facts of Canada's growth and development are so many and varied that there can be no doubt about them But what of the future? Far be it from me to don the robe of the prophet. However, it might be useful at this point to summarize some of the facts on which the future will be built and on which we base our hopes

First, there is the fact of our natural resources.

Our forests, farm, fishing and trapping industries with research and planning can be organized on a perpetual yield basis. Sustained yield logging and improved silviculture techniques in our forests, improved methods of cultivation and new strains in cereals, restocking of fishing waters, conservation of wild life and general emphasis on conservation could ensure the use of these resources in perpetuity.

As to foodstuffs, we shall continue to be an exporter of high quality wheat and other grains, and of the products of our fisheries but as population increases there will be less of other foods to export.

Stored in the Canadian Shield are the raw materials of many metals which are an indispensable foundation for our machine age The developments in the Precambrian rock of this shield with its many intrusions already have been spectacular, and much of this area is practically unexplored It would be amazing if the future does not bring many more discoveries just as spectacular.

The Paley Report says "The United States has crossed the great industrial divide and from being a nation with a surplus of raw materials has become a deficit nation." If what this report says is true, the United States will be increasingly interested as time passes in the riches of the Canadian Shield and in other mineralized areas of Canada.

There is an important corollary to the discovery of new mineral areas. The opening of the iron mines of Labrador-Quebec included the building of 360 miles of railway. The new mining town at Lynn Lake in Northern Manitoba involved the building of a railroad of 155 miles. All such developments mean a further opening up of the northern areas. Great reserves of lead and zinc have been discovered at Pine Point on Great Slave Lake, which is 800 miles from the International

Border. Eventually this territory also is likely to be reached by railroad. So the east-west development of Canada is now being supplemented by one in depth.

These resources of the forest and mine are not a sufficient basis for confidence in the future in themselves. We would be condemned to be hewers of wood and drawers of water for more fortunately situated countries if we did not possess another indispensable resource for a country with ambition to become highly industrialized; that is, abundant supplies of energy. Low cost hydro-electric power has been fundamental in the development of our industries. For example, in our pulp and paper industry a combination of forest resources, rushing streams to carry down the logs and furnish hydro-electric power, has given us a leading position.

The only reason we can have an aluminum industry is the existence of abundant low cost hydro-electric power. Kitimat is an excellent example Our metallurgical and electro-chemical industries are dependent on this same source of energy.

Our supply of power has increased steadily. The present turbine installation of 15 million horse power represents less than a quarter of our economic potential of nearly 66 million. This does not include the potential at the Grand Falls on the Hamilton River in the Quebec-Labrador iron ore area New developments are tending to make it practicable to transmit power economically over much larger distances, and the harnessing of this great falls is under discussion. The potential is estimated as of the order of 10 million horse power. It is almost impossible to keep up to date in these developments. Only recently it has been announced that preliminary talks were under way to promote a Canadian plan for a giant hydro-electric development which would harness the Yukon river and provide for a new industrial area in the far northwest. Full utilization of the water flow available in the Yukon and northern British Columbia could, eventually, produce 4 to 5 million horse power.

But hydro-electric power may not be enough for our future needs even when augmented by that from the St. Lawrence Waterways and other developments. It could be supplemented by the coal of the Maritimes, the Prairies, and British Columbia, by the oil and gas carried by pipe lines from the Prairies, and eventually by the use of atomic energy from the uranium of which Canada is a leading producer

The fact is that Canada is in the process of being bound together by a veritable network of equipment for the transmission of power sources —hydro-electric transmission lines, oil and gas pipe lines.

More than natural resources and power is necessary for the successful

development of the Canada of the future. We need more population; but even there we are making progress. Our population has been increasing in recent years at a rate of about 2½% per annum and is now over 15,000,000. It is a very pessimistic person in Canada who does not visualize a population of between two or three times our present one at the end of the century. Taking current birth and death rates and current net immigration and assuming they continue at the same rates, we would have a population of 18 or 19 million in 1961, 22½ million in 1971.

Of course, also, the growth and development of Canada will continue to require a vast amount of capital investment. Oil and gas can be discovered and brought to the surface, mines developed, industries built up, pipe lines and railroads constructed, towns erected, only with the aid of capital. These investments in their development and subsequent operation provide the work opportunities for a growing population.

Have we good reasons for thinking the capital will be forthcoming to build the future Canada? Let us look at the record. While Canadians have financed by far the largest part of post-war investment, non-resident capital has been an important source of financing some major developments in recent years. At present other nations have $11.2 billion worth of capital invested in Canada. Of this amount $8.6 billion is owned in the United States compared with $4.1 billion in 1939. From 1945 to 1953 the United States increased its investments in Canada by $3.6 billion.

Americans have $3.6 billion of portfolio investments in Canada which pay interest or dividends but involve no control of the enterprise. Direct investments amount to $5 billion which are wholly owned or in which there is majority control in the United States.

These direct investments in branch plants or subsidiary or controlled organizations are another instance of the benefits which our neighbor has bestowed on Canada. They have accelerated the pace of development of our natural resources and hastened our industrialization. They are advantageous from our point of view because when in operation in many cases they supply the means of repayment. A mine or pulp and paper mill produces commodities which can be exported to the United States. Also, successful direct investments usually mean a ploughing back of profits in the Canadian branch or subsidiary. It is a permanent type of investment and multiplies. Such operations reduce the net demand for U.S. dollars on the part of the Canadian economy and, therefore, relieve strains on our balance of payments. One of the greatest benefits, of course, is that along with the creation of the branch plant or subsidiary come the technical skills and knowledge accumulated through years of experience by the parent concern.

Canada, therefore, has welcomed these large American capital investments, and their welcome has been all the heartier because there has been a growing tendency to train and bring Canadians into the management and also to permit Canadians to participate in the financing. On the whole these branches have adapted themselves to the Canadian atmosphere and have respected Canadian aims and co-operated in their attainment. Except in name, in actual operation they are scarcely distinguishable from Canadian concerns.

Of course, I would not have you think that this large American financial interest in Canada means that your countrymen own the bulk of it. Even in the four industries in which American capital is mostly concentrated—manufacturing, mining and smelting, petroleum exploration and development, and public utilities, Canadian ownership increased from 62% in 1939 to 68% in 1951, and the proportion owned by Canadians of all the national wealth is very much greater than this

All things considered, lack of capital should not be an obstacle to the growth of Canada In fact, Canada is not now heavily dependent on outside sources for capital Approximately 80% of the post-war expansion was financed by Canadians

Our ability to keep our trade, not only commodity but also invisible items such as interest, dividends, tourist trade, and a host of others in balance, is another factor on which future development will depend This problem, of course, is known technically as the Balance of International Payments.

During the year 1953 the debits in our current account with the United States exceeded our credits by $924 million We had a credit balance with all other countries of $485 million which gave us an overall debit of $439 million Failing any other source of meeting this adverse balance, it would have been necessary to draw on our national reserve of United States dollars But the change in this reserve was not significant The explanation of that phenomenon is the flow of United States and other funds into Canada for investment, which offset the adverse balance on current account and created such a demand for Canadian dollars that the latter has been at a premium over the United States dollar

Suppose, however, there was a slowing up of this capital investment, what would be the situation if a large deficit remained in our current account? Is it to be expected that year after year this huge investment of capital by United States corporations and individuals will continue without at least periods of slowing up? This brings us face to face with the question of trade relationships.

Exports are the life blood of the Canadian economy. The products of our farms, mines, forests, and fisheries are far beyond the capacity of

our small population to consume. The importance of foreign trade is evident in the fact that in 1953 we ranked third in both exports and imports among the trading nations of the world, surpassed only by the U.S.A. and U K.

In view of the importance of foreign trade to Canada, our government, throughout the post-war period, has worked for the reduction of existing barriers to international trade. Canada is one of the few countries in the world which has almost no significant barriers to imports other than tariffs, and the Canadian tariff has been reduced considerably since the war. Incidentally, Canada has never received directly one dollar of Lend Lease or Marshall Plan aid Instead, she has contributed substantially in mutual aid, loans and other forms of international assistance like the Colombo Plan.

Our needs have set Canada in the very forefront of international efforts for freer trade. This policy has been pursued in the face of discouraging obstacles. Many of the obstructions to overseas trade were created by the imposition of quotas and embargos so that the mutual tariff concessions negotiated in the General Agreement on Tariffs and Trade have yielded little benefit to Canada. Canada has lived up to the agreements but overseas countries could not because with them tariff was not the difficulty but the fact that they were short of dollars

The United States adoption of the Reciprocal Trade Agreement Program and some other supporting developments, such as the original sponsoring of the General Agreement on Tariffs and Trade, were landmarks in this field. Recently, however, progress has been retarded, and in certain cases there has even been a movement in reverse

There has been disappointment that simplification of your customs procedures, particularly in respect to the system of classification and valuation, has not progressed much farther than is the case. Resumption of the discussions next spring of these trade barriers, in many cases more protective than tariffs, will be followed here and abroad with keen interest.

In spite of barriers of which we wish there were fewer, there is an enormous flow of goods across the border. The United States is our best customer. In 1953, 59% of our exports were to the U.S.A. The other side of the picture is that we are a very good customer since we purchased 73.5% of our imports from your country in 1953. In monetary terms our exports to the United States amounted to approximately \$2½ billion and our imports from the United States to approximately \$3¼ billion. Since the war we have imported annually, on the average, over \$500 million worth of goods more than we have sold to your country. Incidentally, 56% of our exports to the United States in 1953 were raw or partly manufactured materials. The other 44%, i.e., fully

processed materials and finished goods includes the important item, newsprint, in which the value added by manufacture is low. On the other hand, 84% of our imports from the United States were fully processed materials and finished goods.

How does this trade situation affect our hopes for the future? There are a number of factors which will have a bearing on it.

First, the question of trade is, of course, part of the wider subject—the balance of international payment. When one nation invests heavily in another, a large part of the investment usually takes the form of a flow of commodities If the investment diminishes, the commodity flow lessens Thus the adverse balance of trade on the receiving end would be reduced.

Second, hopes are rising that those countries which have been short of dollars are nearing the stage where convertibility of currencies will be achieved—a necessary basis for freer world trade.

Third, a great deal of the development which has been going on in Canada has yet to reach the production stage Its fruition should affect our trade picture markedly. It is difficult to be pessimistic about the future of Canadian trade when one thinks of the aluminum which is to come out of Kitimat, the iron ore from Ungava, the higher production of nickel and other non-ferrous metals from numerous new mines, titanium from the Allard Lake region when the process is working commercially, and the possibility of another new metallurgical colossus in the northwest through the harnessing of the mighty Yukon. Canadians look to the development of the St Lawrence Waterways as another means of strengthening their economic position.

There will be, of course, increased offsets to earnings from exports by transfers of profits on U S. capital investments, but, at the same time, a number of developments will diminish our need for imports. We shall arrive at self sufficiency in oil, and gas flowing through the Trans Canada gas pipe line to the east will replace some imported coal.

The growth of our population will mean a reduction in the amount of foodstuffs we shall have for export On the other hand, Canadian growth and development and our rapidly expanding economy will produce opportunities for the domestic production of manufactured or semi-manufactured goods which have been imported in the past. This is true already in our domestic iron and steel industry and in a variety of chemical products.

(It is interesting to note that in the period 1926–1929 exports of goods and services from Canada accounted for some 29% of gross national expenditure, in 1936–39 the proportion was about 28%, and in 1950–53 about 23%. The growth in Canada's population and the domestic market seems slowly to be reducing the proportion of

Canada's resources devoted to direct production for the foreign market.)

Nevertheless, Trade could be our Achilles Heel. Canada is dependent on foreign trade to a much greater extent than the United States with its huge domestic market. Canada will thrive more than most nations in an era of peace and growing prosperity. In a depressed economic climate or dollar shortage she will have her problems. With sanity and vision on the part of labour and management in relation to costs; given a world in which the forces of peace gain the upper hand and in which progress towards the ideal of raising world standards of living can be made, the economic future of Canada surely should be very bright.

Some of the lesser developed nations surpass Canada in resources of some minerals, or in forest areas, or in power potentials. Canada has no rival among them in the combination of mineral and forest resources, energy potentials, a belief in free enterprise, the stability of sound government, and a safe harbor for capital investment. Canada undoubtedly has an impressive array of assets and I am sure the United States would like to see our country continue its progress so that this whole North American continent may present the strongest front to authoritarian aggressiveness.

Man shall not live by bread alone. Canada might have great material prosperity and yet be in danger of losing her own soul. At least, this is the warning which many eminent Canadians have been voicing in recent years. Our concern about cultural influences led to the appointment in 1949 of a Royal Commission on National Development in the Arts, Letters and Sciences which produced a masterly report. It has become known as the Massey Report after its Chairman, who is now the first Canadian to be Governor General of the Dominion.

It emphasizes the fact that Canada owes a great debt to the United States for the generous manner in which your cultural facilities have been made available to Canadians. The gifts of the Carnegie Corporation and the Rockefeller Foundation; the aid of the Guggenheim Foundation and the American Association for the Advancement of Science to a host of students seeking opportunities for further study; the fellowships and scholarships awarded to Canadian students in American Universities, tell a tale of magnificent generosity.

Moreover, our nearness to the United States and many similarities in our ways of life mean that a vast mass of other cultural media pour across the border. These include radio, films, newspapers, periodicals, books, and your touring musicians and artists. Although in recent years Canadian periodicals have strengthened their position greatly, they have had to face almost overwhelming competition. Canadians read more American periodicals than they do Canadian,

with the exception of local newspapers.

All this influence, however, has not been an unmixed blessing for the development of Canadian culture. Educational opportunities made available to Canadians has resulted in a loss of much talent since many students, when they finish their courses in American institutions, accept positions in the United States and do not return to Canada.

Perhaps we have been too dependent on American generosity. The brief of the National Conference of Canadian Universities to the Royal Commission commented "American generosity has blinded our eyes to our necessities Culturally we have feasted on the bounty of our neighbors, and then we ask plaintively what is wrong with our progress in the Arts."

Here I might mention that since the Report of the Royal Commission was made public the Dominion Government has instituted a system of special grants to Universities.

In Canada, as in the United States, the battle between the fundamentalist and progressivist philosophies of education has been joined. The need for more education in the humanities has been given increasing emphasis Quite recently Hilda Neatby, a professor in one of our Canadian Universities, wrote a book severely critical of the present situation Entitled "So Little for the Mind" it became a best-seller in the non-fiction field and has aroused a storm of discussion which may be the forerunner of reform.

Canadians, of course, can tune in on American radio programs and have as their movie film diet mainly the Hollywood output Many of your radio programs are superlative in character and afford immense enjoyment to numerous Canadian listeners There are other aspects of American broadcasting which have been severely criticized in your own country. The Massey Report urged "That in Canada measures should be taken to avoid in our Radio and T V. at least those aspects of American broadcasting which have provoked in the United States the most outspoken and the sharpest opposition "

Our Canadian Broadcasting system is the result of recommendations made by a Royal Commission on Radio, known as the Aird Commission appointed in 1928 The view has been accepted that radio is a Public Trust and not just another industry; its social influence is so great that for the welfare of its citizens government must take some responsibility in the matter of control. In the Canadian Broadcasting Act of 1936 the Canadian Broadcasting Corporation (C B.C) was created and given wide powers of control Broadcasting was to become a force fostering a national spirit and interpreting national citizenship. The C.B.C. policy is to concentrate Canadian resources on the development of programs essentially Canadian and to supplement them with the best

available entertainment from other countries, which means mainly the United States.

The C.B.C. exercises control by virtue of the fact that it recommends to the Minister of Transport the grant, renewal or cancellation of licenses to private operators. Also, it can regulate the nature and amounts of advertising, political broadcasts, and generally, the character of all programs whether public or private.

While, of course, there is room for improvement in the operations of C.B.C., the Massey Report states that it "has opened the way to mutual knowledge and understanding which would have seemed impossible a few years before. Canadians as a people have listened to news of their own country and of the world, have heard public topics discussed by national authorities, have listened to and participated in discussions of Canadian problems, and have, through radio, been present at great national events." In short, C.B.C. has been of immense importance in developing our cultural identity.

Such is the system concerning radio in Canada. Since television also comes under the jurisdiction of the C.B.C., its development may be expected to follow the same pattern.

Time will not permit me to enlarge on this subject. While it is true that mediocre movies, the juke box, soap operas, comic papers, jazz and so on occupy too prominent a place in our cultural atmosphere, it is also true that there is a widespread and growing interest in music, drama, painting and the fine arts generally. In the last two summers there was a record-breaking movement of visitors from all directions to witness the Shakespearian Drama Festival at Stratford, Ontario. Canadian artists have won acclaim abroad and Canadian musical composers were thrilled recently by the success of the all-Canadian concert at Carnegie Hall under the leadership of Leopold Stokowski. As time passes Canadians will make distinctive contributions to the culture of North America.

In conclusion, we Canadians do not wish to become a pale and insipid copy of the great American people. We wish to have a distinctive Canadian way of life which, while benefiting by your great qualities, will evolve also from a fusion of our own rich French, Anglo-Saxon and other heritages, developed and refined in the crucible of Canadian experience. We must think for ourselves and stand on our own feet. Sometimes, therefore, we may have the temerity to disagree with you! With the fundamentals of your way of life there will be complete accord. The northern neighbor will be one with you in promoting and defending the freedoms which have been inherited and enhanced, and in all efforts to secure peace, goodwill, and higher standards of living for the human race.

THE POPULATION OF THE UNITED STATES IN 1950 CLASSIFIED BY AGE, SEX, AND COLOR—A REVISION OF CENSUS FIGURES*

ANSLEY J. COALE
Princeton University

IT IS clear to all who consider the question that data from the United States decennial censuses of population are less than perfect. Persons who should be included are omitted, others are counted twice, and characteristics of the persons included are sometimes misreported. This article is addressed to errors of omission and mistaken inclusion in the 1950 census, and to the erroneous classification of persons according to their age, sex, and color. No mention will be made of census data relating to kinship, employment status, occupation, education, income, etc.

We will survey the evidence that reveals imperfections in the census, analyze some of the evidence, and offer a set of numbers which we believe come closer than census figures to the United States population in 1950, classified by age, sex, and color

EVIDENCE OF IMPERFECTIONS IN CENSUS DATA

Anyone familiar with the scale of effort required to complete a census of population and with the necessarily limited resources made available for the job should not be surprised that the enumeration is less than completely exact. The fact that some 140,000 inexperienced enumerators and field supervisors serving with moderate pay and no prospect of prolonged employment conducted the actual house-to-house canvass leads naturally to the expectation that not every person in every district was properly counted No amount of care and planning by the Census Bureau could insure perfection on the part of all the members of a large army of temporary employees Similarly, the best designed questionnaire and enumerator's instructions could not guarantee that the respondent would give and the enumerator record accurate information about all the members of a given household.

* Many of the data on which this article is based were supplied by the Bureau of the Census from unpublished material The author wishes to express his appreciation to the Census Bureau for making this material available In addition to giving the author access to unpublished data, members of the Bureau's staff were very generous in giving advice and comments at various stages in the preparation of this work Among those who were especially helpful are Henry S Shryock, Jr , Henry D. Sheldon, and Richard A Hornseth of the Population and Housing Division; and Joseph F. Daly and Leon Pritzker of the Office of the Assistant Director for Statistical Standards. At the same time that their help is acknowledged, it must be made clear that neither the Bureau of the Census nor any of the members of its staff in any way accepts, approves, or endorses the estimates herewith presented.

Internal Evidence

Though the nature of the census leads us to expect imperfect data, an examination—even an exhaustive one—of census procedures would not give any quantitative indication of errors. To obtain clues as to how *many* persons are missed and how *many* are misclassified requires a comparison of census numbers with other numbers. Actually, a limited amount of information about errors can even be obtained from a comparison of census numbers with other numbers from the *same census.*

Two informative kinds of "internal" comparisons are *sex ratios* and *age ratios* [7]. Sex ratios are formed by dividing the number of males at any age by the number of females; age ratios are formed by dividing the number of persons in any five-year age group by the average of the numbers in the two adjacent age groups. These two kinds of ratios will reveal census errors because in each instance we can estimate roughly what the ratios *ought* to be.

As for sex ratios, one would generally expect the ratio of males to females in a closed population to exceed unity among young children, and to decline continuously as age advances. This expectation is based on the almost universal tendency for male live births to exceed female by some 3 to 6 per cent, and for female mortality to be less than male at every age. The native white and native nonwhite populations are presumably a near approximation to "closed" populations, particularly when the "population abroad" is included. Figure 1 shows sex ratios by age in 1950 for native whites and nonwhites, along with the ratios one would expect on the basis of the sex ratio at birth and the survival rates (taken from life tables) to which each cohort has been subject[1] In neither instance do the ratios follow the values one would expect. The sex ratios among young adults are particularly far from expected values. Nonwhite sex ratios deviate from expected values by more than 10 per

[1] The "expected" sex ratios were calculated by taking appropriate values from United States life tables as far back as these have been prepared (to 1920), and life tables (for the white population) prepared for the Death Registration Area back to 1900. Thus to obtain the expected sex ratio at ages 40–44, we multiplied the probability of surviving from 30–34 to 40–44 in 1940–1950 (taken from the 1945 life table) times the probability of surviving from 20–24 to 30–34 in 1930–1940, times the probability of surviving from 10–14 to 20–24 in 1920–1930, times the probability of surviving from 0–4 to 10–14 in 1910–1920 (this last estimated by averaging 1909–11 and 1919–21 life table values). This procedure was followed separately for males and females. The resulting male value was multiplied by 1.055 for whites and 1 03 for nonwhites (representing the sex ratio at birth) and divided by the female value. The life table mortality experience of the male cohorts aged 20–24 through 45–49 was corrected for estimated excess mortality due to military deaths in World War II. Though this procedure may be inexact (e.g., because it employs life tables for whites instead of native whites and for the Death Registration Area instead of all states), the fact is that sex ratios are not very sensitive to changes in mortality experience. Hence, except for errors in the life tables themselves caused, for example, by differential reporting of male and female deaths, the "standard" sex ratios should be a reliable guide.

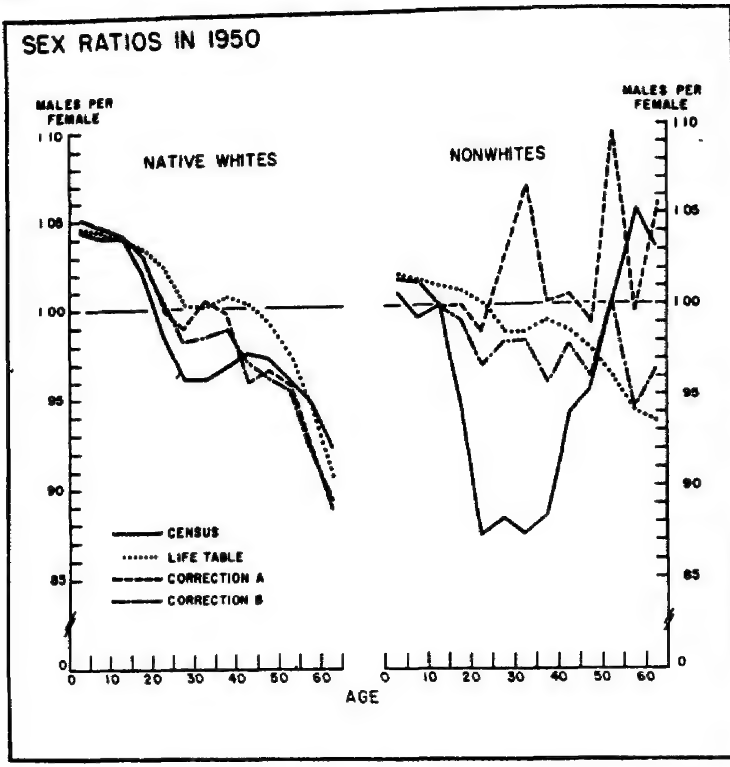


FIG 1. Males per female by five-year age groups Native whites and nonwhites (including persons living abroad) according to the 1950 census and according to a "cohort" life table

[Life table data from which cohort life tables were derived are from U S life tables for 1945, 1930–39, 1920–29, an average of tables for 1909–11 and 1919–21, and 1901–1910 The 1945 life table was published in. U S. National Office of Vital Statistics, *Vital Statistics—Special Reports*, Vol 23 (1947) The 1930–39 life table was published in U S Bureau of the Census, *U S Abridged Life Tables, 1930–39*. (Preliminary.) The earlier life tables are in U S Bureau of the Census, *U. S Life Tables, 1930*. The war-caused excess mortality was calculated from figures given in U S National Office of Vital Statistics, *Vital Statistics of the U S., 1949*. Part I.]

cent at ages 20 through 39, and to a similar degree (though in the opposite direction) at ages 55 through 64.

A roughly similar pattern of sex ratios has characterized earlier censuses In Figure 2 the sex ratios from the last three censuses for native whites and for nonwhites are shown In all three censuses the young adult ages have too low a sex ratio, while the sex ratio for the

older ages is clearly too high in all instances except among native whites in 1950. There is a marked tendency for the sex ratio at older ages to progress toward more reasonable values from 1930 to 1950. The differential reduction in mortality rates (with male rates reduced more than female) would account for only a small part of the reduction in census sex ratios.

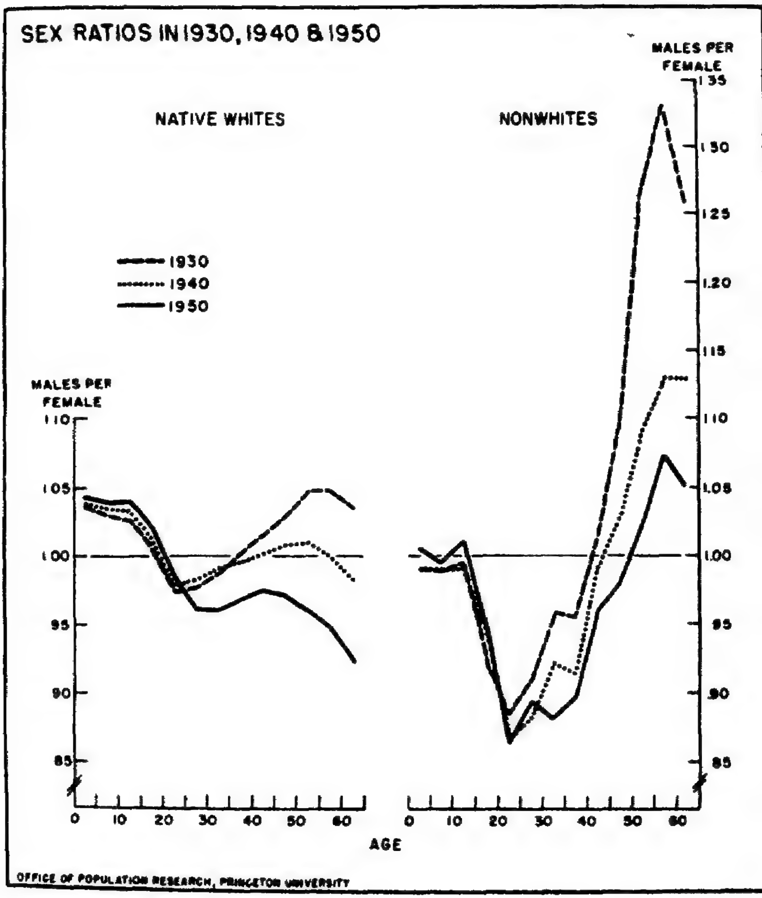


FIG 2. Males per female by five-year age groups according to the 1930, 1940, and 1950 censuses. Native whites and the nonwhite population. 1950 ratios for whites include the population living abroad.

The calculation of what the age ratios ought to be is a little more complicated. Crudely speaking, age ratios ought to be unity, on the assumption that numbers in adjacent age groups progress smoothly, and do not depart markedly in a short interval from a straight line. We may distinguish three factors, however, that would cause age ratios to depart from unity. The first factor is the tendency (if any) of typical mortality experience to produce a strongly non-linear age distribution. This factor seems to be unimportant up to very advanced ages [7]. The second factor is any temporary variation in birth or migration rates which produces an unusually large cohort. The third factor is the tendency for ages to be reported so that the proportion of persons in some age groups is overstated and the proportion in others understated. This tendency might be called age-heaping, and "five-year age-heaping" when we are dealing with five-year age groups.

Since we are concerned here with census errors, we are interested in the last of these factors—that of five-year age-heaping. The first factor may be assumed to have negligible importance, and the second factor—unusual cohort size—we will try to eliminate We eliminate the effect of unusual cohort size by dividing each age ratio by an adjusted average of the age ratios for the same cohort in every census in which it has been enumerated. Thus if the cohort aged 40–44 in 1950 were unusually large (relative to the adjacent cohorts) and consequently had a large age ratio, one would normally expect the age ratio for the group 30-34 in 1940, for the group 20–24 in 1930, and 10–14 in 1920 to be above unity. Since the cohort is enumerated at a different age in each census, there should be a tendency for "age-heaping" to cancel out of the average age ratio for the cohort. Thus one would expect a given age ratio for a particular census divided by the average ratio for the cohort to differ from unity primarily because of age-heaping in that census.

The reader will have to bear with us through one more stage of complication. We said above that there *should* be a tendency for age-heaping to cancel out of the average for a cohort enumerated in several censuses However, if ages in the first half of each age decade (i.e., ages 10–14, 20–24, etc.) are generally preferred in age reporting, the average age ratio for every alternate cohort would tend to be above unity, while the average for those in between would be below one. An examination of average age ratios for cohorts indeed shows a general alternation of large and small values It was therefore decided to adjust the average cohort ratios for the typical average age-heaping to which a cohort is subject at the particular ages where it was enumerated. Thus the average age ratio for the cohort aged 30–34 in 1950 (which was 20–24 in 1940, and 10–14 in 1930) was divided by the average age ratio of all groups aged

10–14, 20–24, and 30–34 in 1930, 1940, and 1950. Since this last average covers five different cohorts, it will not be dominated by any unusually large or small cohort. If this average differs from unity, the difference arises from the average tendency of the ages 10–14, 20–24, and 30–34 to be favored or avoided. Division by this average adjusts the cohort age ratio to make it more nearly reflect cohort size alone.

In Figure 3 corrected age ratios for the 1930, 1940, and 1950 censuses are charted. The following features are noteworthy:

1. There is a strong similarity of the pattern of age-heaping from census to census, among both sexes, and among whites and nonwhites.

2. Age-heaping is much more pronounced among nonwhites than among the white population. The nonwhite age ratios differ from unity by as much as 20 per cent, while the largest divergence for the white population is less than 5 per cent.

3. There is a tendency up to age 40 to prefer the last five years of each age decade, and to avoid the first five years. This tendency *could* be the result of a reluctance to pass certain 10-year birthdays. On the other hand, the ratio at ages 50–54 is consistently high, and at ages 55–59 consistently low.

4. There is a tendency, especially clear-cut among the white population, for the ratios at any one age to progress in the same direction from 1930 to 1940 to 1950. Usually this progression is toward unity—especially where the age ratio departs radically from 1.0 in 1930. In other words, five-year age-heaping, while occurring at about the same ages, has generally been diminishing.

5 There is one noteworthy exception to the similarity among the age-ratio patterns in 1930, 1940, and 1950. This exception is the departure of age ratios for ages 60–64 and 65–69 in 1940 and 1950 from the 1930 values, especially among the nonwhites, to a lesser degree among white females, and not perceptibly among white males. The character of the change is that the 60–64 group became smaller relative to neighboring age groups, while 65–69 became much larger. This change is from a pattern which extends well back of the 1930 census. The change in age-ratio pattern confirms a shift in age reporting which has been often noticed and commented on.[3] The unexpectedly large number of persons 65–69 and the unexpectedly small number 60–64 have been attributed to the old-age assistance legislation which came into effect between 1930 and 1940. The minimum age of 65 necessary to qualify for benefits may have caused persons for whom otherwise an age of 60–64 would have been reported to be reported as 65–69. The

[3] U. S. Bureau of the Census. *U. S. Census of Population: 1940*, Vol. IV. *Characteristics by Age*, p. 3; and *U. S. Life Tables and Actuarial Tables, 1939–41*, pp. 110–112.

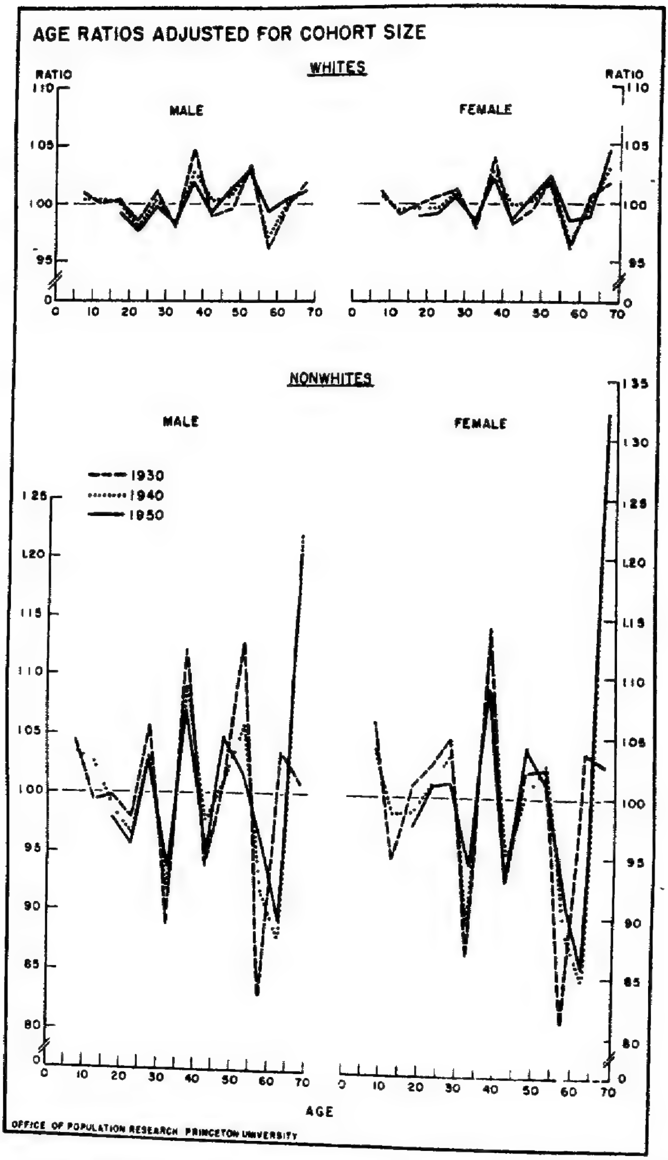


FIG 3 Population in each five-year age group divided by average of populations in adjacent age groups for whites and nonwhites, 1930, 1940, and 1950, adjusted for cohort size by dividing each age ratio by corrected average ratio for cohort

change since 1930 is greatest among nonwhite females, nearly as great among nonwhite males, much smaller but still detectable among white females, and not apparent among white males. These differences suggest that a nonwhite as opposed to a white, and a female as opposed to a male, is most likely to apply for old-age assistance when under 65.

The existence of heaping in the single-year age distribution of the population is another internal indication that misclassifications occur. By Myers' blended method [4], one can calculate the fraction for whom an age ending in each digit is reported in a way that would yield almost exactly 10 per cent for each digit if age were accurately reported. Table 1 shows the apparent preference as indicated by this method for digits of age in United States censuses from 1880 to 1950, and in 1950 for the nonwhite population as well as the total. For the whole population in 1950, the proportion with reported ages ending in zero was about 12 per cent larger than it should have been; for the nonwhite population, the proportion was apparently about 32 per cent too large. Note that improper digit choice has steadily declined since 1880.

TABLE 1
PER CENT OF POPULATION WITH EACH TERMINAL DIGIT 0–9 USED TO REPORT AGE IN U. S. CENSUSES, 1880–1950

Digit	Total population (whites and nonwhites)								Whites	Nonwhites
	1880	1890	1900	1910	1920	1930	1940	1950	1950	1950
0	16.8	15.1	13.2	13.2	12.4	12.3	11.6	11.2	10.9	13.2
1	6.7	7.4	8.3	7.7	8.0	8.0	8.5	8.9	9.0	7.3
2	9.4	9.7	9.8	10.2	10.2	10.3	10.4	10.2	10.2	9.9
3	8.6	9.1	9.3	9.2	9.4	9.4	9.6	9.7	9.8	8.8
4	8.8	9.0	9.5	9.4	9.4	9.6	9.7	9.7	9.8	9.2
5	13.4	12.3	11.3	11.5	11.3	11.2	10.7	10.6	10.5	11.5
6	9.4	9.6	9.4	9.6	9.7	9.6	9.6	9.8	9.8	9.4
7	8.5	8.9	9.3	9.1	9.4	9.3	9.6	9.7	9.8	9.4
8	10.2	10.4	10.2	10.7	10.6	10.5	10.3	10.2	10.1	10.7
9	8.2	8.5	9.7	9.4	9.6	9.8	10.0	10.1	10.1	10.6
Average per cent deviation from 10%	2.08	1.56	.94	1.12	.90	.86	.60	.45	.36	1.20
	(Calculated so that unbiased use of terminal digits would yield 10% for each digit)									

Sources

From 1880 to 1930 Myers, Robert J., "Errors and bias in the reporting of ages in census data." *Transactions of the Actuarial Society of America*, Vol 41 (Part 2), No 104, 1940, pp 395–415.

1940 U S Bureau of the Census. *Sixteenth Census of the United States, 1940. U. S. Life Tables, 1939–41*, p 121.

1950: Computed from U S. Bureau of the Census *U. S. Census of Population 1950*. Vol II, Table 94.

External Evidence

Any close estimate of the numbers of persons omitted from the 1950 census and the numbers with specified misreporting of characteristics requires a comparison with information derived from some outside source. Among the other sources which have been examined in attempting to appraise the 1950 census (or earlier censuses) are [1]:

1. Estimates of the number of young persons derived from the births registered in the years before the census, and from the deaths among young children recorded between the date of birth and the census period.

2. Estimates of persons over 10 years of age derived from earlier censuses, and from death and international migration statistics for the intercensal interval. As will be seen later it is primarily by an extension of this method that the age-sex-color distribution for 1950 offered below is obtained.

3. Draft registration of males of military age. The selective service registration of 1940, which was a mandatory registration for all draft-age males in the United States (with certain exceptions, such as those already in the armed services) in October, 1940, provides numbers which, after adjustment for deaths, migration, and differences in coverage, can be compared with the 1940 census

4. The Post Enumeration Survey conducted in 1950. This survey provided more carefully determined numbers for a representative sample of areas in the United States. From the discrepancies observed in these areas, it is possible to estimate what the nationwide discrepancies would have been had the Post Enumeration Survey's careful enumerative methods been used nationally One of the important features of the Post Enumeration Survey is that it permits a name-by-name check on part of the census population rather than merely a comparison of totals. This feature is particularly important to an investigation of the reporting of personal characteristics by the census.

We will leave to one side for the moment all comparisons except those derived from birth data and other censuses. The differences, with regard to persons classified as white males, between numbers derived from this source and numbers from the 1950 census are shown in Figure 4.

Differences between the expected and census population in 1940 and 1950 form remarkably similar patterns. The pattern alone does not tell us much, however, about shortages and overcounts in the censuses, since an apparent undercount can be caused (for example) by an actual undercount, by an overcount of the same cohort in the earlier

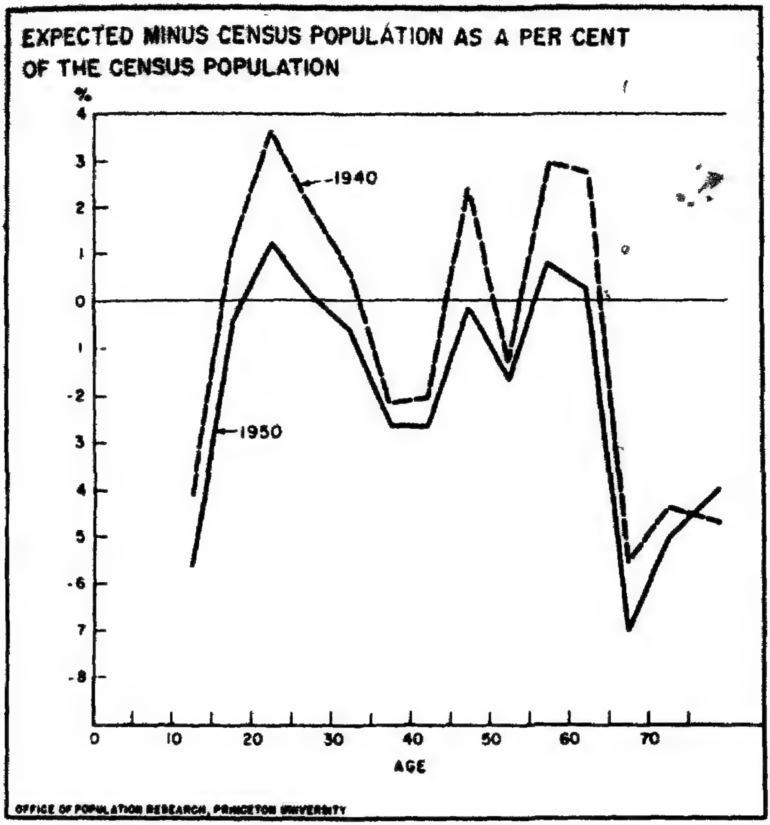


FIG. 4. Differences between the population expected on the basis of the preceding census and estimated cohort changes and the enumerated population, as a per cent of enumerated population, 1940 and 1950, white males.

census; or by an underestimate of the intercensal decline in the size of the cohort.

The pattern in Figure 4 can be explained in part by the age-heaping we have already discussed The excess of expected overenumerated population at 45–49 is as one would expect—since, judging from Figure 3, 35–39 seems to be a more strongly preferred age group than 45–49. But the very low level of the curves at 10–14 and the relatively high level from 15–34 would *not* be expected on the basis of age-heaping evidence. On the other hand, if enumerated white males under five are compared with the number expected on the basis of registered births

(corrected for underregistration), an apparent undercount of 7 per cent in 1940 and about 4 5 per cent in 1950 is revealed Thus the deficiency of expected compared to the enumerated population at ages 10–14 may reasonably be attributed to a relatively large undercount of the cohort in an earlier census.

THE HYPOTHESIS THAT RECENT CENSUSES ARE SUBJECT TO SIMILAR ERRORS

Much of the evidence we have surveyed so far would support the supposition that the 1930, 1940, and 1950 censuses were characterized by errors which form similar patterns By similar patterns we mean that the variation of undercounts and overcounts according to age, sex, and color are alike from one census to the next

Some of the factors lending credence to this hypothesis are:

1. Many features of census taking which would be most likely to affect omissions, erroneous inclusions, and misclassification by age, sex, and color did not change substantially. Persons were enumerated at their places of regular residence in all three censuses; all three enumerations required a period of several weeks to complete; in all instances the enumerators were temporary employees selected by similar procedures, the questions were generally asked of one respondent for each household in all three censuses, age was recorded in response to essentially the same question ("How old was he on his last birthday?") and in no instance in response to a question about date of birth; etc. There were *some* procedural changes that may well have altered the pattern of errors somewhat The introduction of special infant cards in 1940 (for children under four months of age) may have had a tendency to reduce the undercounting of the very young; the new rules in 1950 designating the college rather than the home as the usual residence of college students may have affected errors in counting college-age groups; and the "T-Night" procedure in 1950 may have caught transients who would have been missed by the techniques used in earlier censuses. Nevertheless, the general similarity of census-taking methods would support the belief that if young adult males were "missed" in 1950, they were probably missed in 1940 and 1930 as well; and that if ages were misreported in a particular way in one census, they were similarly misreported in the others

2 The pattern of age ratios by sex and color, and sex ratios by age and color showed striking similarities in the three censuses. (See Figures 2 and 3) As to age ratios, the most notable *shifts* of pattern (at ages 60–64 and 65–69) are explainable by new legislation. The sex ratios in

all censuses are too low among young adults and too high among persons 40–64. There is, however, an observable tendency for extreme age ratios to progress toward unity, and for sex ratios among older persons to move toward more reasonable values.

3. Finally, when populations expected on the basis of earlier censuses and estimated cohort changes are compared with enumerated populations, the strongly similar pattern of discrepancies suggests that omissions and misstatements are similar from one census to the next.

We propose to use the hypothesis of similar errors as a means for forming new estimates of overcounts and undercounts in the 1950 census. The hypothesis is the basis for new estimates when combined with estimated intercensal cohort changes, and with independently estimated errors for special age groups in one census One determines the error in 1950 for certain age groups—say 5–9 and 10–14—by an independent method, estimates the errors in an earlier census at the same ages by means of the hypothesis of similarity, corrects the earlier census population, and finally forms new estimates of the expected population 15–19 and 20–24 by adding intercensal cohort changes to the revised figures in the earlier census The repeated use of this procedure will generate estimates of errors at all ages. The essence of the procedure is very simple—an assumption of similarity enables us to estimate errors in earlier censuses at the *same* ages for which errors are calculated in a later census; while cohort changes combined with known errors in an earlier census permit us to estimate errors at *later* ages in the later census.

SYSTEMATIC COMBINATIONS OF ESTIMATED COHORT CHANGES AND THE HYPOTHESIS OF SIMILAR ERRORS

We will use several variants of the hypothesis that error patterns were similar in the last three censuses, and will note the resulting differences in estimated undercounts and overcounts. Later on we will try to take account of the previously noted differences among the censuses with respect to five-year age-heaping

Specifically, the following procedures are employed:

1 The population in 1950 at ages 0–4, 5–9, and 10–14 by color and sex is estimated from registered births (corrected for underregistration), deaths, and migration. These estimates are deemed correct, and differences between them and the 1950 census are interpreted as census undercounts.

2. The 1930 and 1940 censuses are assumed subject to the same per cent undercounts at ages 5–9 and 10–14 as the 1950 census.

3. Estimates of intercensal cohort changes (made by the Bureau of the Census from registered deaths and from migration statistics) are added to the estimated numbers at age 5–9 and 10–14 in 1930 and 1940 to obtain numbers at ages 15–19 and 20–24 in 1940, and 15–19, 20–24, 25–29, and 30–34 in 1950 The census undercounts or overcounts at these ages are then calculated.

4. It is assumed that (*a*) the 1930 census was subject to the same per cent undercount or overcount as the average of the 1940 and 1950 censuses, or (*b*) that the 1930 census was subject to an undercount the same as that in 1940 or 1950, whichever was smaller.

5. These assumptions permit the calculation of a corrected population at ages 15–19 and 20–24 in 1930; and by adding cohort changes, one can estimate the numbers 25–29 and 30–34 in 1940, and 35–39 and 40–44 in 1950 By an iterative process it is possible to continue the calculations to the oldest ages

6 Another variant using the assumption of similarity in the pattern of errors is to calculate undercounts at ages 0–14 in 1950 as above, and to assume for ages above five that the 1940 census was subject to the same per cent under- and overcounts as the 1950.

Table 2 presents undercounts and overcounts by age, color, and sex in 1940 and 1950 according to three different assumptions of similarity in error patterns In the first columns of Table 2 it is assumed for ages above 15 that per cent over- or undercounts in 1930 were the same as in 1940 or 1950, whichever was less; in the middle columns average of 1940 and 1950 miscounts are assumed for 1930; while in the last columns 1940 miscounts are assumed equal to those of 1950.

A MORE REFINED HYPOTHESIS OF SIMILAR ERROR PATTERNS IN THE 1930, 1940, AND 1950 CENSUSES

The three versions of the hypothesis of similar error patterns used as the basis for Table 2 take no account of the differences among the last three censuses in five-year age-heaping which are evident in Figure 3. A more refined version of the hypothesis would assume that the errors in one census were the same as in another, after explicit allowance for observed changes in five-year age-heaping. The reason such refinement is worthwhile is that it produces a more logically consistent set of age-by-age estimates of errors in 1950 Figure 3 suggests, for example, that ages 35–39 were more strongly preferred in 1930 than in 1940 or 1950. Indeed our rough calculations (briefly described in Appendix I) yield estimates that 3 8 per cent fewer nonwhite males were heaped into ages 35–39 in 1950 than in 1930. On the basis of this estimate, one would

TABLE 2

ESTIMATED ERRORS IN THE 1940 AND 1950 CENSUSES AS A PER CENT OF THE CENSUS POPULATION BY SEX, COLOR, AND 5-YEAR AGE GROUPS ACCORDING TO 3 DIFFERENT ASSUMPTIONS ABOUT SIMILAR PATTERNS OF ERROR

Age	Assuming errors in 1930 equal to the average of 1940 and 1950 errors				Assuming errors in 1930 equal to those in 1940 or 1950, whichever is less				Assuming errors in 1940 equal to those in 1950			
	WM	WF	NWM	NWF	WM	WF	NWM	NWF	WM	WF	NWM	NWF
						1950						
0–4	4.5	3.8	11	10	4.5	3.8	11	10	4.5	3.8	11	10
5–9	3.1	2.5	12	10	3.1	2.5	12	10	3.1	2.5	12	10
10–14	1.1	1.1	7	7	1.1	1.1	7	7	1.1	1.1	7	7
15–19	2.6	1.5	18	12	2.6	1.5	18	12	2.6	1.5	18	12
20–24	2.3	.9	19	6	2.3	.9	19	6	2.3	.9	19	6
25–29	4.4	1.5	27	9	4.4	1.5	27	9	2.8	8	27	11
30–34	4.1	−.3	27	4	4.1	−.3	27	4	1.7	.3	32	10
35–39	2.8	−.6	20	7	1.9	−1.0	20	5	.1	−1.3	25	7
40–44	1.4	2.1	21	14	.2	1.8	18	10	−1.0	−.5	26	7
45–49	2.6	3.1	23	17	2.1	2.7	20	16	0.0	.6	27	15
50–54	.4	1.0	14	8	−.9	−.5	11	1	−2.7	−.5	25	9
55–59	6.7	7.8	36	44	5.8	8.4	33	40	8	3.6	38	32
60–64	.7	4.9	25	30	−1.0	2.5	18	15	−3.0	1.9	42	19
65–69	1.7	2.2	21	5	−.6	.9	15	0	−6.0	−3.0	13	−10
70–74	−.5	2.3	44	15	−2.3	3	34	−1	−9.6	5	40	—
75+	−7.5	−6.2	73	34	−11.1	−9.1	47	20	−24.9	−11.3	113	48
Total	2.6	1.8	19	11	2.0	1.4	18	9	2	6	22	10
						1940						
0–4	7.1	6.5	19	17	7.1	6.5	19	17	4.5	3.8	11	10
5–9	3.1	2.5	12	10	3.1	2.5	12	10	3.1	2.5	12	10
10–14	1.0	1.1	7	7	1.0	1.1	7	7	1.1	1.1	7	7
15–19	4.1	2.2	18	10	4.1	2.2	18	10	2.6	1.5	18	12
20–24	4.7	.4	15	0	4.7	4	15	0	2.3	9	19	6
25–29	5.5	1.6	22	10	4.6	1.2	22	9	2.8	.8	27	11
30–34	4.1	2.8	27	17	2.9	2.6	24	14	1.7	3	32	10
35–39	3.6	1.1	21	9	2.1	.7	19	8	.1	−1.3	25	7
40–44	1.9	1.0	16	6	.6	−.5	13	0	−1.0	−.5	26	7
45–49	5.2	4.4	25	23	4.4	5.0	23	21	0.0	6	27	15
50–54	.3	2.2	14	17	−1.1	.1	9	7	−2.7	−.5	25	9
55–59	6.9	8.2	46	48	5.1	7.1	41	43	8	3.6	38	32
60–64	3.1	4.0	45	30	1.9	2.5	39	18	−3.0	1.9	42	19
65–69	.9	.7	9	−6	−.8	−.1	4	−10	−6.0	−3.0	13	−10
70–74	−2.7	.7	26	11	−5.4	−1.2	11	−2	−9.6	5	40	—
75+	−3.4	−4.5	10	−9	−5.3	−6.0	−1	−9	−10.5	−7.0	51	29
Total	3.5	2.4	19	12	2.8	1.9	17	10	.8	1.0	23	10

Sources

Changes in cohorts taken from U. S. Bureau of the Census. *Current Population Reports*, Series P-24, No 94, 1954. Estimates of the population 0–4 in 1940 and 0–14 in 1950 based on births adjusted for underregistration, deaths, and migration from the same source. The method of calculation is described in the text.

assume that if nonwhite males were undercounted by $x\%$ in 1950, the undercount in 1930—being compensated by greater age-heaping—would be only $(x-3.8)\%$. The effect in 1950 (because of our technique of using cohort changes) is to reduce the estimated undercount at ages 55–59 by about 3 8 per cent. This more refined hypothesis is used in preparing the figures we finally select as our estimates of the 1950 population. The more acceptable age ratios in the corrected population —evidence that the age-to-age errors are more logically consistent—are shown in Figure 6. More will be said about these age ratios later.[3]

VALIDITY OF THE ERROR ESTIMATES BASED ON THE HYPOTHESIS OF SIMILAR ERROR PATTERNS

The corrections to the 1950 census derived from the hypothesis of similar error patterns may be tested in several ways, none entirely conclusive or satisfactory We can examine critically the steps by which the estimates are constructed, we can show how the estimates would be affected by plausible variations in the basic figures used, we can test the internal consistency of the corrected figures by computing age and sex ratios; and we can compare these corrections with other error estimates—notably those derived from selective service figures for young males in 1940, and from the Post Enumeration Survey in 1950.

Critical Appraisal of the Method Employing the Hypothesis of Similar Errors

The validity of this method of estimating census errors depends, of course, on the truth of the assumption that census errors *are* similar as postulated, and also on the accuracy of the data we have on births and on intercensal cohort changes The evidence underlying the assumption that errors form a similar pattern from census to census has been summarized earlier and will not be re-examined We will, however, examine briefly the quality of the data on which the error estimates rest.

The number of live births in the United States from 1935 to 1950 plays a key role in this method of estimating census errors. Estimates of the "correct" population up to age 15 depend directly on these birth data, and a 1 per cent error in estimating the number of births would cause a corresponding error in estimating the corrected population under 15 Moreover, because of the iterative character of the method, a 1 per cent change in birth figures would bring about a very nearly equal change in the corrected population at *all* ages.

The birth figures employed are the number of registered live births,

[3] See pp. 35–39 below.

corrected for underregistration. It is fortunate that birth registration has been subject to check on two occasions so that convincing estimates of underregistration are available Estimates of underregistration are based on the results of matching individual registrations against individual census records in 1940 and 1950. From these matchings the degree of underregistration in each match period (four months in 1940 and three months in 1950) was estimated separately by state of birth, color, and according to whether the birth occurred in a hospital or elsewhere Corrections to registered births between 1940 and 1950 were based on the assumption that completeness of registration for each category of birth changed linearly between the two tests. For the years 1935–1939 underregistration within each category was assumed constant at the 1940 level.[4]

The third support on which this method rests—in addition to the assumption of similar error patterns and the accuracy of birth data—is the quality of the estimates of intercensal changes in cohort size The estimates of cohort changes we have used were prepared by the Bureau of the Census [10].

Cohorts change in size because of death and international migration. The estimates of cohort changes from 1930–1940 and 1940–1950 were based on somewhat different methods For the earlier decade the fraction of each cohort surviving was calculated from the U S. Life Table for the decade. Net international migration was assumed to be zero for every cohort (This assumption rests on the fact that the official data on international movements for the *total* population show only a few thousand net migrants during the intercensal period) For the interval between the 1940 and 1950 censuses, cohort changes due to mortality were estimated from annual registered deaths Estimated decedents were subtracted from each cohort, whereas from 1930 to 1940, cohort attrition was estimated by means of a survival ratio Cohort changes due to migration in the later decade were estimated from data supplied to the Census Bureau by the Immigration and Naturalization Service. These data included information on age, color, and sex only for *alien* migrants, the age, color, and sex of *citizen* migrants (including migrants to and from U. S. possessions) had to be determined for the most part by guesswork

The only allowance made for underregistration of deaths was the assumption that deaths to children under one are underregistered by the same per cent as births No test for the completeness of death regis-

[4] The matching procedure has been described in some detail [2, 5, 6], and the estimates used here of the corrected number of births for each year have been published in an official source [11].

tration comparable to the tests of birth registration has been attempted. However, it is generally believed among public health statisticians (on the basis of the plausible thesis that a physician is more likely to be in attendance or a public official to be notified in the event of a death than in the event of a birth) that deaths are more completely registered than births [12].

On the other hand, the survival rates used for the 1930–1940 period depend on population figures as well as on registered deaths, and this dependence introduces an additional possibility of error in the estimates of cohort change Moreover, even if death registration were sufficiently complete for our purposes, erroneous reporting of the age of decedent, and reporting of his color which is not consistent with census classification, could be the source of false estimates of the change in specific cohorts.

There is a widespread and credible impression that international migration statistics are deficient. A recent instance serving as a source of skepticism is the illegal and presumably unrecorded entry of large numbers of "wetbacks" from Mexico. There are other notable deficiencies in the migration portion of the estimated intercensal cohort changes in addition to unrecorded entries and exits. The negligible level of total migration between 1930 and 1940—if the over-all figures are accepted as valid—does not *really* imply that there were no significant gains or losses among particular cohorts It is more likely that there were slight gains among the younger adult cohorts and slight losses among the older ones Also the estimation of the age, color, and sex of non-alien migrants between 1940 and 1950 is little more than a guess.

The Effect of Plausible Variations in the Data on the Estimated Corrections

The effect of variations in the assumption of a similar error pattern is illustrated in Table 2. The identical value of the estimated errors under age 25 in Table 2 for all three assumptions merely reflects the fact that in each instance errors at ages 5–9 and 10–14 are assumed equal to those in 1950, and errors under 15 in 1950 are derived from birth data.

The significance of the differences in Table 2 will be more apparent if we first consider what would be the effect on this method of estimating errors if one census differed slightly from others at *all* ages. Suppose, to be specific, that the 1930 census in addition to having the characteristic errors of age-heaping, and of omission in particular age groups (such as young children) had more or less consistently omitted 1 per cent *more* persons at each age than the later censuses. The result would

be that the "expected" population in 1940 and 1950 would be too small starting with the earliest ages dependent on the assumption of similar errors; and hence the undercount estimated at these ages would be about 1 per cent too small (or the overcount 1 per cent too large). These mistaken estimates of errors in 1940 and 1950 would cause a correction 2 per cent too small to be applied to the corresponding age groups in 1930. The net effect would be estimates of a purportedly corrected population in 1950 which would be 1 per cent too small at young adult ages, 2 per cent too small twenty years older, and 3 per cent too small forty years older.

If only two censuses are involved—as is the case in the third part of Table 2—the cumulative effect of a consistent 1 per cent deficiency in the earlier census is more rapid and pronounced. Thus if the 1940 census were 1 per cent less fully enumerated at all ages than the 1950 census, the "correct" population at ages 15–24 would be 1 per cent underestimated; at 25–34, 2 per cent underestimated; at 35–44, 3 per cent; etc.

The foregoing feature of our method (that a systematic difference in censuses produces errors which cumulate at the older ages) accounts for many of the differences in Table 2. Thus one would expect the assumption of average rather than minimal errors in 1930 to produce a larger corrected population in 1950, with the differences becoming greater with advancing age Also the relatively small undercount at 25–44 and the rapidly rising indicated overcount at the upper ages among white males in the third part of the table suggests that white males were generally subjected to a smaller count in 1940 than in 1950. This impression is confirmed by Figure 4 which shows the population expected on the basis of the preceding census minus the enumerated population expressed as a per cent of the latter. This difference is uniformly lower in 1950 than in 1940, as would be the case if white males were consistently subjected to a small count in 1940 as compared to 1950 and 1930 As a result we will discard (at least for the white population) the assumption that errors in the 1940 census were the same as in 1950, and restrict our choice to some form of assumption about the 1930 census. The assumption of smaller errors in 1930 is conservative, though the fact that it leads to estimates of substantial overcounts above age 65 may mean that it is too conservative with regard to the white population.

* * *

Defects in our basic data would affect the validity of the estimated errors in varying degree. We have already pointed out that a given per cent mistake in the estimated completeness of birth registration will

produce an error of equivalent magnitude in the corrected census population. However, the estimates of birth underregistration seem well grounded.

The effect of deficiencies in the mortality data would be much less, because cohort attrition due to death is less than 10 per cent up to age 50 among whites and up to age 40 among nonwhites. Thus an underregistration of deaths by 10 per cent would cause less than a 1 per cent error in estimating the size of a cohort under those ages. Table 3 shows what would be the cumulative effect among nonwhite males on estimated errors in 1950 if deaths are assumed to be 5 per cent more numerous than the registered figures. If the general belief that deaths are more fully registered than births is correct, our estimated errors are not subject to serious question because of death registration.

TABLE 3

APPARENT UNDERCOUNTS IN THE 1950 CENSUS AS A PER CENT OF THE CENSUS POPULATION FOR NONWHITE MALES, ASSUMING THAT THE 1930 CENSUS WAS SUBJECT TO THE PER CENT ERROR IN 1940 OR 1950, WHICHEVER WAS LESS, AND 5 PER CENT UNDERREGISTRATION OF DEATHS

	0–4	5–9	10–14	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–59	60–64	65+
No Underregistration	11	12	7	18	19	27	27	20	18	20	11	33	18	29
5% Underregistration	11	11	6	18	19	26	25	18	16	18	7	28	11	18

The magnitude of errors introduced into our estimates by defective migration data can only be guessed at Unrecorded *immigration* would produce an *under*estimate of the correct population; unrecorded *emigration* an *over*estimate Some impression of the effect of plausible variations in the estimation of cohort change arising primarily from different treatment of migration data can be derived from Table 4. In both parts of Table 4 errors in the 1950 census are calculated from the same assumption about error patterns—namely, that the 1930 census was subject to the lesser of the errors in 1940 and 1950. The two sets of estimates differ in that they depend on different estimates of cohort changes between 1940 and 1950 Both estimates were derived by the Bureau of the Census.[6] The mortality figures used were the same in both instances (except for a different estimate of military deaths in World War II) The principal differences arise from different estimates

[6] One set of estimates is from U S Bureau of the Census. *Current Population Reports*, Series P-25, No 98, August 1954. The other is derived by interpolation between an estimate for July 1, 1949 and July 1, 1950. *Current Population Reports*, Series P-25, No 39, and *Statistical Abstract of the U. S.*, 1951, p. 10

of net migration, including different assumptions about the age, sex, and color of non-alien migrants Another source of difference is that the first estimates of cohort change were compiled with the benefit of more complete information relating to the armed services overseas in 1950.

TABLE 4

ESTIMATED ERRORS IN THE 1950 CENSUS AS A PER CENT OF THE CENSUS POPULATION, BY SEX, COLOR, AND 5-YEAR AGE GROUPS ACCORDING TO TWO ESTIMATES OF COHORT CHANGES BETWEEN 1940 AND 1950

Age	Cohort changes from *Current Population Reports*, Series P-25, No 98				Cohort changes interpolated from *Current Population Reports*, Series P-25, No 39 and *Statistical Abstract of the U S*, 1951			
	WM	WF	NWM	NWF	WM	WF	NWM	NWF
0–4	4.5	3.8	11	10	4.5	3.8	11	10
5–9	3.1	2.5	12	10	3.1	2.5	12	10
10–14	1.1	1.1	7	7	1.1	1.1	7	7
15–19	2.6	1.5	18	12	2.4	1.5	18	12
20–24	2.3	.9	19	6	3.0	1.1	17	5
25–29	4.4	1.5	27	9	3.5	1.3	20	8
30–34	4.1	−.3	27	4	2.4	0.0	16	4
35–39	1.9	−1.0	20	5	1.3	−.8	14	5
40–44	.2	1.8	18	10	1.5	1.9	15	10
45–49	2.1	2.7	20	16	1.8	2.9	16	15
50–54	−.9	−.5	11	1	−.7	−.1	1	1
55–59	5.8	8.4	33	40	5.6	8.5	25	40
60–64	−1.0	2.5	18	15	−.6	2.8	7	14
65–69	−.6	.9	15	0	−.4	1.2	16	4
70–74	−2.3	.3	34	−1	−1.9	−.2	33	10
75+	−11.1	−9.1	47	20	−13.7	−10.3	0	−1
Total	2.0	1.4	18	9	1.8	1.5	14	9

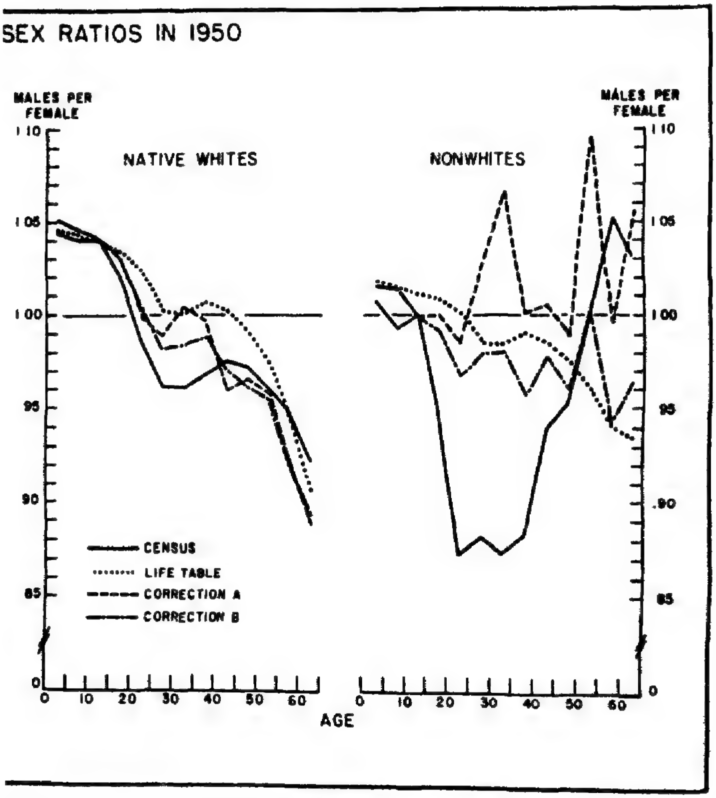
In both sets of estimates it is assumed that errors in the 1930 census were equal to those of 1940 or 1950, whichever was less.

Sex and Age Ratios for the Corrected Population

In Figures 1 and 3 we subjected the 1950 census to tests of internal consistency and found that sex and age ratios deviated from plausible values We will now put populations corrected by the method of similar errors and cohort change to the same tests. We will test several different sets of corrections, which will be labeled A, A', B, and B'. Each set of corrections is derived from the assumption that the 1930 census was subject to errors equal to the smaller of the errors in 1940 and 1950. The differences between the primed and unprimed corrections is that the primed corrections allow for changes in age-heaping, while the unprimed do not. The A corrections use the estimated cohort changes for 1940–1950 published in 1954, the B corrections use earlier estimates of

›hort change. The following diagram summarizes the different bases · the four sets of corrections:

	No allowance for age-heaping changes	Allows for age-heaping changes
epends on cohort changes for 1940–50 ›rived from CPR, Series P-25, No. 98	*A*	*A'*
epends on cohort changes for 1940–50 ›rived from CPR, Series P-25, No 39	*B*	*B'*



ᶠIG. 5. Males per female by five-year age groups, native whites and nonwhites ›luding persons living abroad) according to the 1950 Census of Population, ording to corrections A and B, and according to a cohort life table.

Figure 5 shows sex ratios among the native white and nonwhi populations in 1950 according to the census as calculated from tables and according to corrections A and B.[6] Among the white popul tion either correction A or correction B produces a more plausible s of sex ratios from age 15 to age 40 than does the census. Above age 4 the corrected sex ratios seem consistently too low. Among nonwhit the sex ratios resulting from the B corrections are a striking improv ment over the census, and are also much less erratic than the sex rati resulting from the A corrections.[7] This superiority leads us to prefer t B corrections for the nonwhite population.

Figure 6a shows age ratios for 1950, divided by adjusted cohort a ratios for the census, and for corrections A and A', while Figure shows the same ratios for corrections B and B'. The significant fact be observed in this figure is that a crude assumption of similar errors 1930—i e , either corrections A or B—leads to age ratios which are le credible than those in the census itself. The age ratios in the populati corrected by methods A or B deviate from unity just as far as in t census and exhibit a regular sawtooth pattern which shows the existen of some kind of systematic error. However, when either set of corre tions is altered by assuming that the 1930 census was subject to t error in a later census modified by calculated differences in age-heapin a set of age ratios closer to unity and essentially free of a regular sa tooth character is produced [8]

Among white males the best age ratios result from the B' rather th the A' corrections, while among nonwhite males the A' corrections sult in slightly better age ratios

In all instances but one, the age ratios resulting from the refin hypothesis of similar errors (the primed corrections) are a clear i provement on the census. This exception is method A' among the whi males. However, for white males the B' corrections give age rat which are typically closer to unity than census ratios.

---

[6] In preparing Figure 5 it was assumed that the native white and native nonwhite populations w subject to the same undercounts and overcounts as the total white and nonwhite populations.

[7] The very high sex ratios at ages 25–29 and 30–34 in the A corrections are the result of the la number of nonwhite male net immigrants estimated for those cohorts in the *Current Population Repo* Series P-25, No. 98. Net immigration between 1940 and 1950 of nonwhite males who were 30–34 1950 is estimated as 69 thousand. There are figures in the 1950 census which cast some doubt on plausibility of immigration this extensive The sum, in the 1950 census, of foreign-born nonwhites, nonwhite Puerto Ricans, plus Filipinos, at ages 30–34 (which would include the major portion of nonwhite immigrants *whenever* they entered the country) is less than 15 thousand.

[8] It should be noted that it is an adjustment to the 1930 census correction at age 35–39 which the effect of producing sensible age ratios at age 55–59 in the corrected population, and also that adjustments are calculated from census data alone without reference to the corrections themselves. other words, there was nothing in the adjustments employed in methods A' and B' which automatica insured an improvement in age ratios.

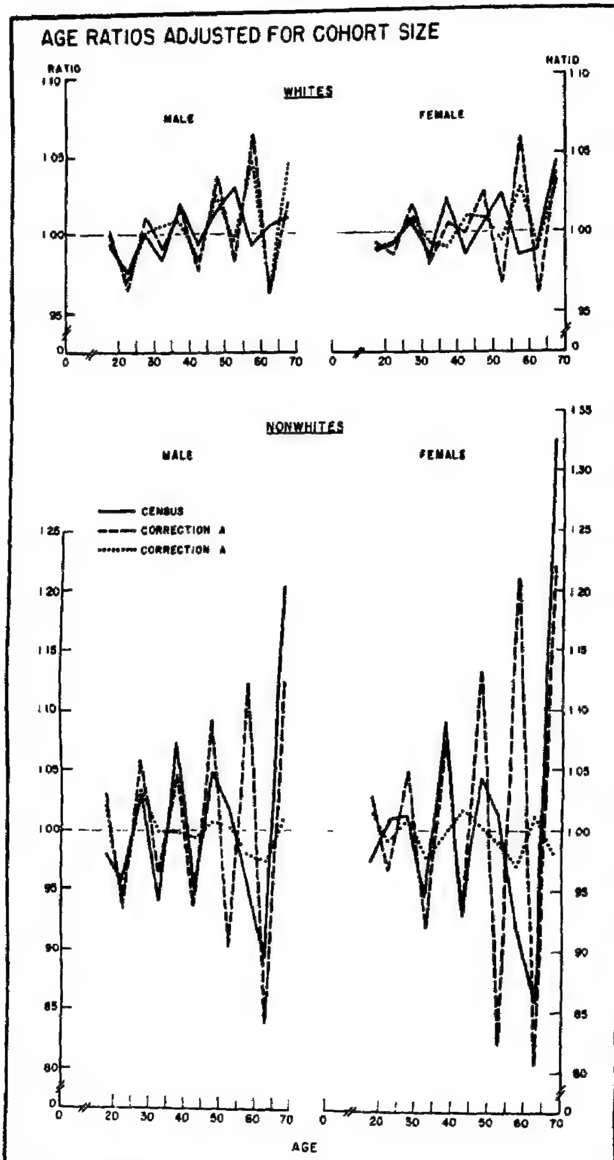


FIG. 6a. Age ratios in 1950 divided by adjusted age ratios for each cohort for the white and nonwhite populations, according to the Census of Population, and according to corrections A and A'.

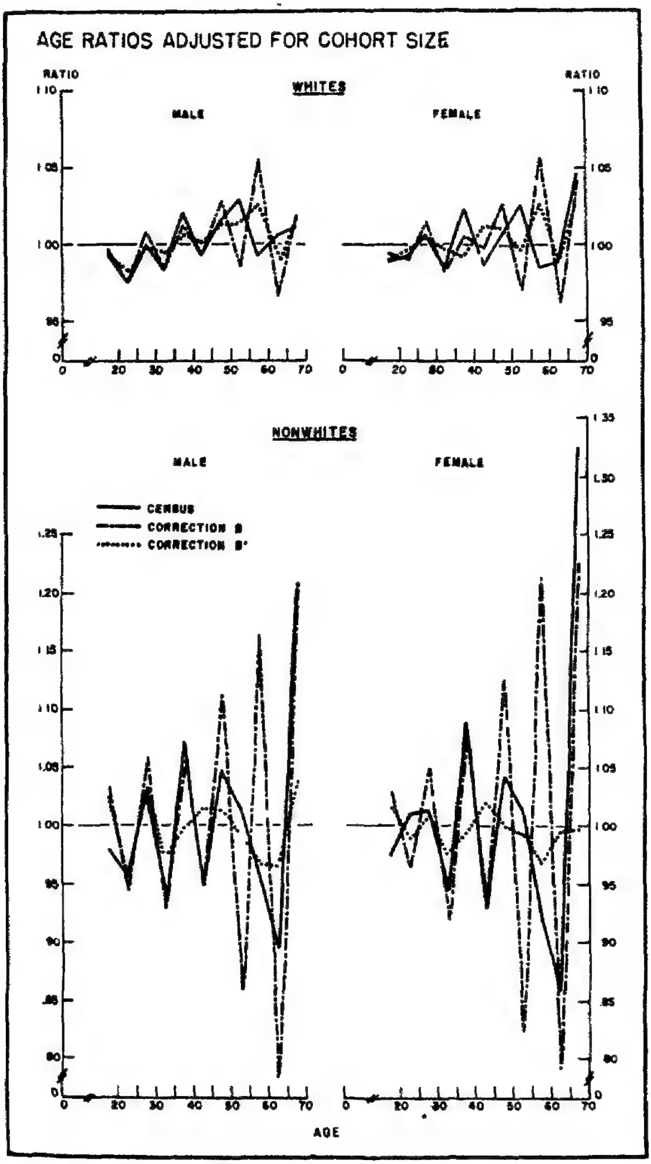


FIG. 6b. Age ratios in 1950 divided by adjusted age ratios for each cohort for the white and nonwhite populations, according to the Census of Population, and according to corrections B and B'.

Comparisons of Census Errors Estimated by the Method of Similar Patterns with Other Estimates

Table 5 shows the undercount of males of draft age in 1940 (21–35) compared with the undercount calculated for males 20–34 by various versions of the method of similar error patterns.

TABLE 5

UNDERCOUNT AMONG MALES OF DRAFT AGE IN 1940 BY COLOR AS A PER CENT OF CENSUS POPULATION ACCORDING TO:

Color	Selective Service figures	Various versions of similar error hypothesis					
		Errors in 1930 equal to average of 1940 and 1950	Errors in 1940 equal to those in 1950	Errors in 1930 equal to 1940 or 1950, whichever is less			
				A	*A′*	*B*	*B′*
White	4.5	4.8	2.3	4.1	4.9	4.3	5.1
Nonwhite	18	21	26	20	23	20	22
Total	5.8	6.4	4.7	5.7	6.7	5.8	6.8

Sources Estimated errors in the 1940 census for males 21–35 on the basis of Selective Service data taken from A Ross Eckler and Leon Pritzker, "Measuring the accuracy of enumerative surveys" (paper presented at the 27th Session of the International Statistical Institute, New Delhi, 1951). U. S Bureau of the Census (processed), p. 2.

Any of the methods using an assumption of similar error patterns with the exception of that which assumes 1940 and 1950 to be alike shows errors of roughly the same magnitude as do the selective service data This comparison serves to reinforce the plausibility of a large undercount among young adult males Young males are among the most mobile members of our society and the census methods of enumeration may well be particularly deficient in counting mobile persons.

In Table 6 errors based on various versions of the similar error method are compared with errors derived from the Post Enumeration Survey. With regard to the white population, our methods yield much larger errors at the younger ages than does the PES, while at the oldest ages the PES indicates an undercount in contrast to the apparent overcount shown by the other methods. Among nonwhites our methods yield estimated errors uniformly larger than those of the PES.

It is clear from Table 6 that our methods and the Post Enumeration Survey can scarcely be considered as mutually confirming. The differences between them can for the most part be explained, however, by a conjecture to the effect that persons of certain types who were missed by the census were also missed in the Survey.[9] This conjecture seems plausible enough in view of the similarities between census and Survey procedures. Although the Survey involved a careful recanvass of a sample of areas to obtain an estimate of households completely omitted, plus a reinterviewing in a sample of households to estimate the erroneous inclusions and omissions within households; although the Survey was conducted only by enumerators chosen after aptitude and performance tests from among the regular census enumerators; and although these enumerators were given intensive supplementary training, nevertheless it is readily imaginable that, for example, persons who are mobile and relatively rootless would be missed by both canvasses. The regular Post Enumeration Survey did not attempt to cover the population residing in dwelling places occupied by more than 35 persons;[10] moreover, because the Survey was conducted 4 to 6 months after the census, it could hardly serve as an adequate check on transients. And if there were any tendency for nonwhite respondents to be reluctant (for example) to give information to strangers about (in particular) male members of the household, such reluctance would doubtless be as strong in a Post Enumeration Survey as in a census. In other words, the Post Enumeration Survey might be expected to reveal many (though not all) of the errors which could be attributed to poor training and gross carelessness on the part of the enumerators, and also many of the errors in reporting personal characteristics which are a result of obtaining information from a respondent other than the person enumerated. However, many errors caused by faulty memory or deliberate concealment, and omissions caused by such factors as impermanent residence would *not* be revealed by the Post Enumeration Survey.

THE ESTIMATION OF ERRORS AT THE OLDER AGES

34a

The least reliable results of our method of estimating errors are at the older ages. These results may be distrusted on two grounds: first,

[9] This possibility is suggested on page 7 in the introduction to Volume II of the 1950 Census of Population [9].

[10] There was a "T-Night" coverage check on the enumeration of places patronized by certain transients conducted by the Bureau of the Census, but it could not parallel the regular Post Enumeration Survey procedure. The results of the coverage check have not been published, but it appears from preliminary figures that the check, when completed, would not more than begin to account for the discrepancies in Table 5.

TABLE 6

APPARENT ERRORS IN THE 1950 CENSUS AS A PER CENT OF THE CENSUS POPULATION, BY COLOR, SEX, AND BROAD AGE GROUPS, ACCORDING TO VARIOUS METHODS OF ESTIMATING ERRORS

Method of estimating error	Color and sex	Age 0–24	Age 25–44	Age 45–64	Age 65+	Total
The Post Enumeration Survey	White					
	M	.9	.2	2.5	2.0	1.1
	F	.5	.7	3.1	1.8	1.2
	Nonwhite					
	M	3.1	3.9	0.1	12	3.3
	F	4.5	4	4.6	4.6	3.3
Version of Similar Error Hypothesis						
Errors in 1930 equal to average of 1940 and 1950 errors	White					
	M	2.9	3.2	2.6	−1.7	2.6
	F	2.1	.7	4.0	−.6	1.8
	Nonwhite					
	M	13	24	24	41	19
	F	9	8	22	15	11
Errors in 1940 equal to those in 1950	White					
	M	2.9	1.0	−1.2	−12.8	.2
	F	2.1	−.1	1.3	−4.8	.6
	Nonwhite					
	M	13	28	31	47	22
	F	9	9	17	9	10
Errors in 1930 equal to those in 1940 or 1950, whichever is less						
A	White					
	M	2.9	2.8	1.6	−4.3	2.0
	F	2.1	.5	3.1	−2.6	1.4
	Nonwhite					
	M	13	23	20	29	18
	F	9	7	17	6	9
A'	White					
	M	3.0	3.4	1.5	−4.6	2.2
	F	2.2	1.1	4.0	−2.2	1.8
	Nonwhite					
	M	13	25	22	29	19
	F	10	9	21	10	11
B	White					
	M	3.0	2.2	1.6	−4.8	1.8
	F	2.1	.6	3.4	−3.0	1.5
	Nonwhite					
	M	12	17	12	16	14
	F	9	7	16	4	9
B'	White					
	M	3.1	2.8	2.3	−2.5	2.4
	F	2.3	1.1	4.2	−2.3	1.9
	Nonwhite					
	M	13	18	15	16	15
	F	10	9	21	8	11

Source: Errors from the Post Enumeration Survey are adapted from preliminary unpublished results of the Survey. The final census report on the Survey has not been published. The final figures the Census Bureau derives from the Survey may differ from these unofficial data.

because some of the results themselves do not seem plausible, and second, because the method is inherently weakest at the older ages.

The features of the estimates for older persons which seem least credible are:

1. The apparent overcount of over 2 per cent for whites over 65.

2. The sex ratio for nonwhites over 65 which ranges from .945 for method B' to .986 for method A, whereas the sex ratio over 65 derived from life-table values is only .857.

The overcount among whites over 65 seems unreasonable in part because the PES shows an undercount of about 2 per cent among older whites. While one can readily understand that the PES would err by missing persons also missed in the census, it is hard to see how the Survey would find a fictitious undercount.

The inherent weakness at the older ages of our method of estimating errors is that the recursive technique employed leads to the possibility of cumulating at the older ages the errors introduced by spurious data or inaccurate assumptions Thus the systematically greater deficits among whites we have detected in the 1940 census lead to especially bad results in the older ages. Also a fixed underregistration of deaths would have its greatest effect on estimates for the older ages

In Appendix II another possible explanation of the deficiencies in the older age error estimates is explored This explanation is that the completeness of enumeration at the ages above 50 has improved since 1930. Our method assumes that the 1930 census was just as complete as the 1950 census; if in fact the 1930 census was *less* complete at older ages, the correct count in 1950 would be underestimated by our method.

Because of the weaknesses in the error estimates for older ages by our method, we will in our final estimate of the errors in the 1950 census use the Post Enumeration Survey figures for persons over 65.

FINAL ESTIMATE OF ERRORS IN THE 1950 CENSUS

The figures we have selected as the best estimate of the 1950 population are presented in Table 7 For persons under 65, we have taken our figures from method B'. The Post Enumeration Survey results have been used for persons over 65.

These particular corrections have been chosen because they are better in certain details (such as age and sex ratios) than other corrections generated by other versions of our method. We have good reason for rejecting the hypothesis that 1940 errors were at the same level as 1950, but there is no real rationale for assuming minimal errors in 1930. Also—as already stated—inaccurate estimates of cohort change may

have made our corrections wide of the mark. Nevertheless we believe that our estimates give the right order of magnitude to the errors in the 1950 census.

The corrections we have selected indicate that there were about 5 million persons more in the United States in 1950 than were counted by

TABLE 7

THE TOTAL POPULATION RESIDING IN THE U. S. AS OF APRIL 1, 1950, BY AGE, COLOR, AND SEX, ACCORDING TO THE CENSUS AND ACCORDING TO ESTIMATED CORRECTIONS

(In thousands)

Age	White						Nonwhite					
	Males			Females			Males			Females		
	Census count	Corr pop	Error as % of census	Census count	Corr. pop	Error as % of census	Census count	Corr. pop	Error as % of census	Census count	Corr. pop	Error as % of census
0–4	7,244	7,570	4.5	6,940	7,200	3.8	992	1,100	11	987	1,090	10
5–9	5,915	6,100	3.1	5,681	5,820	2.5	799	890	12	804	880	10
10–14	4,945	5,000	1.1	4,750	4,800	1.1	716	760	7	709	760	7
15–19	4,686	4,800	2.4	4,645	4,720	1.5	626	740	18	661	740	12
20–24	5,003	5,190	3.7	5,176	5,270	1.8	604	720	19	699	760	8
25–29	5,350	5,540	3.5	5,575	5,640	1.3	622	750	20	695	750	8
30–34	5,081	5,280	3.8	5,276	5,350	1.3	544	660	20	617	680	9
35–39	4,956	5,030	1.5	5,103	5,050	−1.0	562	630	12	626	650	3
40–44	4,574	4,680	2.3	4,617	4,750	2.9	497	600	20	517	610	18
45–49	4,080	4,140	1.5	4,089	4,200	2.7	446	500	13	455	510	12
50–54	3,756	3,800	1.3	3,779	3,860	2.1	373	420	11	364	420	16
55–59	3,351	3,500	4.6	3,345	3,600	7.6	279	330	17	260	340	30
60–64	2,829	2,880	1.9	2,823	2,980	5.4	208	260	24	198	270	36
65+	5,360	5,470	2.0	6,014	6,120	1.8	437	490	12	459	480	5
Total	67,129	68,980	2.7	67,813	69,360	2.3	7,704	8,830	15	8,051	8,930	11

Grand Total Census 150,697.
Grand Total Corr Pop 156,100.
Error as % of Census Pop. 3.6.

Source Corrections for ages 0–64 are estimated by the method described in the text as B'. For persons over 65, the corrections are adapted from unpublished preliminary results of the Post Enumeration Survey

The corrected population is rounded to the nearest 10,000.

the census The number of males was understated by about 3 million and the number of females by about 2 million The white population was apparently subject to an undercount of nearly 2.5 per cent and the nonwhite population to an undercount between 12 and 13 per cent. The largest understatements of numbers were among young children, young adult males, and persons aged 55–59.

It must be emphasized that the apparent deficits (and overstatements) presented in Table 6 are a result of the combined effects of omission and misclassification. In other words, a table listing apparent deficits for the various age-sex-color groups does *not* tell us characteristics of persons omitted by census takers. It appears, for example, that persons 55–59 were omitted in especially large numbers. A more likely explanation for the large deficit at ages 55–59 is that there is a general tendency to underestimate or overestimate the age of persons 55 to 59 years old. Methods of detecting errors which compare gross totals rather than individual records—the method we have employed is of this sort—fail to reveal directly whether a given deficit arises from omission or misclassification.

The large apparent deficits among nonwhites pose an interesting question: are these deficits the result of omissions of nonwhites or the result of misclassification of some sort? In this instance "misclassification" is perhaps an inexact term, since there are no clearly correct standard categories of white and nonwhite The judgment of race ("RACE, White, Negro, American Indian, Japanese, Chinese, Filipino, Other race—spell out" is the heading on the enumerator's form) in the census was left to the enumerator. He was not instructed to ask the race of the respondent. He was told to assume the race of related persons in the same household to be the same as that of the respondent; and he was instructed to ask about race only for unrelated persons. Hence white persons are persons classified as white by a census enumerator, and nonwhite persons are persons classified as Negro, American Indian, Japanese, Chinese, or Filipino by the enumerator. Thus "misclassification" with respect to color might better be termed "inconsistent classification" of the same person as between one census and another, as between a birth certificate and a census form, or as between the entry made by a Selective Service Board and by a census enumerator.

It is clearly possible for a person to be classed as nonwhite in one census and as white in later censuses. This change might arise because of a reduction in segregated neighborhoods, or because of heavy migration from the South. A person classified by an enumerator as nonwhite while living in a segregated neighborhood might well be classified subsequently as white when living elsewhere. Similarly, persons designated as Negro in the South might later be classified as white by a northern enumerator.

These observations on the possibility that persons may be reclassified from nonwhite to white from one census to another should not be con-

strued as equivalent to observations on the phenomenon known as "passing." Changes in classification in successive censuses would reflect nothing more than different judgments by census enumerators, and have no necessary connection with the classification generally made in the community, or even with the classification each person would make for himself. Our estimate of the error in counting the *total* population is not to any important degree affected by whether or not there have been inconsistencies in color classification from census to census. For if some nonwhite persons were missing because of reclassification, then the population designated "white" in the census must have gained recruits by the same reclassification. Hence for every missing nonwhite that we explain by reclassification, there is a newly missing white that can be explained only by omission.

SIGNIFICANCE OF THE ERRORS

The importance of having a correct enumeration of the population depends, of course, on what use one has in mind for the numbers. It is obviously impractical to attempt a review of *all* of the uses of census population data, and to try to comment on the significance of errors in each use. We will limit our remarks to a few comments on the kinds of conclusions likely to be affected and the kinds that are not.

The error of over 3 per cent in the total number of persons would only seldom have important consequences. People given to referring to 163 million Americans would be more accurate if they spoke (as of the beginning of 1955) of 168 million Americans, but their conclusions would rarely be affected.

A common use of the total number of persons is in deriving *per capita* figures—from birth and death rates to average number of movies attended. Presumably an understatement of the population total by 3 per cent will cause such ratios to be 3 per cent too large. However, in very many instances if the truth were known the unassessed error in the numerator of per capita figures would exceed 3 per cent. We can give one example where the numerator has been subject to verification—namely, the birth rate, where the numerator is the number of births. The preliminary crude birth rate for the United States in 1950—the one listed in such authoritative places as the 1950 *Demographic Yearbook* of the United Nations—is derived from census populations and registered births. The underregistration of births according to the test conducted in conjunction with the 1950 census was a little over 2 per cent [11]

Thus the birth rate is subject to two nearly equal (and, it so happens, compensating) errors [11]

The errors in the population count more apt to be consequential are those for certain groups, particularly when they reach levels several times the error in the total. For example, if the corrections we have derived are proper, and if deaths are equally well registered for males and females, part of the difference between male and female mortality rates is spurious If we adjust the 1950 life table for underenumeration of the base population, the difference in expectation of life at age 5 between white males and white females becomes 5 4 instead of 5 5 years,[12] and the difference in $\mathring{e}_5$ between nonwhite males and females becomes 2 6 instead of 3.4 years.

The most striking errors—and the ones which seem most likely to "make a difference"—are the rates of underenumeration among nonwhites If we interpret these as representing nonwhite persons omitted, we are led to such views as that (for instance) the sections of the country with a high per cent of nonwhites contain a larger fraction of the total population than the census shows. Assuming that nonwhite deaths are as completely registered as white, we would conclude that part of the difference in white and nonwhite death rates—to return to mortality statistics—is spurious The corrected expectation of life at age 5 for nonwhite males is 4.2 years rather than 6 years less than that for white males.

Finally, the existence of errors in the census could lead to inappropri-

[11] Thus ironically enough, the preliminary birth rate based on registered births (and the census population) is probably closer to reality than the birth rate based on adjusted birth figures.

[12] Life table taken from U S National Office of Vital Statistics [10]

The effect on expectation of life of correcting the size of population is small because the large corrections occur at ages (e.g., 20–34) where mortality rates are low It is interesting to note that the differentials which result from correcting the life table for errors in enumeration are closer to the differentials among the industrial policy holders of the Metropolitan Life Insurance Company The mortality experience of the industrial policy holders is no doubt non-representative of the U S population in various ways However, one would expect the *enumeration* of these policy holders by the Metropolitan to be relatively accurate.

DIFFERENCES IN $\mathring{e}_5$

Life table	White females—white males	Nonwhite females—nonwhite males	White males—nonwhite males
U S, 1950	5.5	3.4	6 0
U S, 1950, corrected for underenumeration	5 4	2.6	4 2
Metropolitan Life Insurance Co Industrial Policy Holders, 1950	5.3	2.4	2 6

Data on industrial policy holders are taken from Metropolitan Life Insurance Company [3]

ate "blow-ups" of sample figures in estimating national totals—inappropriate because the total population used is too small, and inappropriate because the relative size of age-sex-color groups is misapprehended. Only infrequently, I suspect, would errors from this source be an important factor compared to sampling variability and the errors of measurement within the sample.

* * *

The results of a discussion which has been rather complicated in places can be summarized in a few sentences.

1 The corrected total of 156 million persons is closer to the number of persons in the United States in April 1950 than the census figure of about 151 million.

2. It is our impression after surveying the evidence that the estimated deficit of about 5 million is, if anything, somewhat conservative. However, if we made less conservative assumptions—assuming average rather than minimal 1940–1950 errors in 1930—the estimated deficit would be increased by no more than 10 per cent.

3. The general pattern of age-by-age errors indicated in Table 6 is probably reliable, though the results seem less trustworthy at the higher ages.

4. There is a possibility that part of the large apparent deficit among the nonwhite population is caused by inconsistencies in classifying persons by color If such were *known* to be the case, the estimated total error would hardly be affected, but the estimated number of nonwhite persons omitted would be reduced, and the estimated number of white persons missed would be increased.

5 The importance of census errors can be judged only when it is known how the figures are used. It seems likely, however, that the differential errors for various subgroups will be more important than the error in the total.

A final comment about the difficulties of improving the accuracy of the census count may be in order. The census of population obtains a large amount of information beyond the mere count of persons. As a general rule, methods which would improve the accuracy of counting would cost more money, and within a fixed budget the higher costs of accuracy would mean less information gathered. It would be costly to extend the training of enumerators and only a moderate improvement would result (the meagerness of the improvement is implied by the small errors discovered by the Post Enumeration Survey). A large reduction in errors would probably require a drastic (and no doubt ex-

pensive) change in procedure, such as adding a thorough *de facto* canvass to the present *de jure* type of census and cross checking the results name by name. The institution and subsequent improvement within the past twenty-five years of nationwide vital statistics have made it possible for the first time to verify census accuracy. With such a verification now possible—this article is a crude example—greater accuracy in determining numbers of persons may become available at moderate cost.

APPENDIX I

ESTIMATING DIFFERENCES IN 5-YEAR AGE-HEAPING IN THE 1930, 1940, AND 1950 CENSUSES

The job at hand is to estimate the surpluses and deficits in five-year age groups in the 1930, 1940, and 1950 censuses attributable to "age-heaping." The general principle of the method we use is to ascertain the ratio of the number in a given age group to the number expected from a "smoothed" age distribution; to ascertain the ratio of the average number in the cohort as enumerated in all censuses to the number expected from a "smoothed" distribution of such averages; and to divide the former ratio by the latter and consider the answer a reflection of age-heaping In other words, an age group deviates from a "smoothed" value, we assume, because of age-heaping and because the cohort itself deviates from a smoothed value The latter factor is minimized by using the ratio of a cohort average in all censuses to the smoothed value of such averages as a divisor.

The actual procedure employed for each sex and color in each census was:

(*a*) Determine the total deviation of each 5-year age group from a smoothed value, by dividing the number in each age group by the three-term moving average centered on the group.

(*b*) Determine the first approximation to the deviation due to cohort size by dividing the average reported size of the cohort in all censuses in which it was enumerated by the three-term moving average of such cohort averages.

(*c*) Since, however, the average reported size of a cohort may deviate from the smoothed value because of persistent age-heaping at every enumeration, the first approximation to the deviation due to cohort size was adjusted as follows:

i. Manufacture a synthetic population consisting of the sum of the persons enumerated in each group in 1930, 1940, and 1950.

ii. Average the numbers in this synthetic population for the ages at which each cohort enumerated in the three censuses were enumerated in *any* census (e g , ages 5–9, 15–19, and 25–29 for the cohort aged 25–29 in 1950).

iii. Divide these average numbers by a three-term moving average to estimate typical divergence from a smoothed value of an average cohort size due to persistent age-heaping.

iv The adjusted estimate of deviation due to cohort size was then obtained by dividing the first approximation (see (*b*) above) by the typical divergence due to persistent heaping (see (*c*) *iii* above)

(*d*) Then the total deviation of each age group (see (*a*) above) was divided by the adjusted estimate of deviation due to cohort size (see (*c*) *iv* above) The result was considered a rough measure of the extent of age-heaping

The differences in age-heaping among the three censuses are shown in Table 8

TABLE 8

DIFFERENCES BETWEEN THE 1950 AND 1930 CENSUSES AND BETWEEN THE 1940 AND 1930 CENSUSES AS TO THE PER CENT ERROR IN REPORTING EACH AGE GROUP CAUSED BY AGE-HEAPING

Age	1950–1930				1940–1930			
	White		Nonwhite		White		Nonwhite	
	M	F	M	F	M	F	M	F
10–14[a]	1 40	1 40	3 60	5 40	70	70	1 80	2 70
15–19[a]	20	− 20	−2.20	−4 00	10	−.10	−1 10	−2 00
20–24	00	− 20	0 00	40	− 20	− 40	80	1.00
25–29	− 30	00	−3 20	−3.60	−.20	.00	−2 10	−1.40
30–34	50	70	5 90	8 10	10	.20	2 40	2 80
35–39	−.80	− 60	−3 80	−4 10	− 60	− 60	−2 30	−1 30
40–44	− 60	− 60	1 70	50	10	20	3 00	1 70
45–49	1 40	1 00	1 10	− 30	50	40	−1 20	−3 00
50–54	−2 90	−1 90	−8 10	−1 60	−1 80	−1 30	−5 50	−0 40

[a] At ages 10–14 and 15–19 the difference between the 1950 and 1930 censuses was assumed equal to twice the difference between the 1940 and 1930 censuses.

APPENDIX II

CHANGES IN THE COUNT OF OLDER PERSONS IN RECENT CENSUSES

A conceivable source of weakness in our estimates of census errors in counting older persons is the possibility that the completeness of

enumeration in 1930 was *not* (as assumed) the same as in 1940 or 1950. As a means of determining whether and to what degree enumeration at the older ages has been changing, we can compare over several decades the fraction of various cohorts that appear on the basis of census data to survive a decade with the fraction apparently surviving on the basis of mortality data The enumerated population aged 60 to 64 in 1930 divided by the enumerated population 50 to 54 in 1920 indicates the apparent fraction of this cohort surviving the 1920's, while the ratio of ${}_5L_{60}$ to ${}_5L_{50}$ from a United States life table for the 1920's gives the same ratio derived from mortality data. But the census apparent survivorship is contaminated, so to speak, by migration; to the degree that a cohort gains or loses by international migration, there *should* be disagreement between census and life table ratios The effect of migration can be mitigated, however, by forming the ratio of the survivorship of two adjacent cohorts Such a ratio should be the same whether derived from census or mortality records, provided that, for example, a cohort aged 50–54 at the beginning of a decade loses or gains through migration the same fraction as the cohort aged 55–59. The ratio of the apparent survivorship of adjacent cohorts derived from successive censuses is divided, in Table 9, by the same ratio derived from mortality data. Thus if we call $({}_5S_x)_c$ the fraction of a cohort aged x to $x+5$ at the beginning of a decade surviving 10 years on the basis of census data, and $({}_5S_x)_m$ the fraction surviving on the basis of mortality data, the ratio R_x presented in Table 9 is given by:

$$R_x = \frac{({}_5S_{x+5})_c}{({}_5S_x)_c} \Big/ \frac{({}_5S_{x+5})_m}{({}_5S_x)_m}$$

This tabulation has a number of revealing features:

1 R_x diverges substantially from 1 00 in the United States, Canada, England and Wales, and the Union of South Africa, but is within 1 per cent of unity in Norway, Sweden, and Switzerland These last countries are known on other grounds to have accurate census data. The fact that R_x is so close to one in these countries, then, indicates that R_x values do indeed reflect the presence of census errors (though of course R_x values other than unity could also be caused by erroneous mortality data)

2 There are certain persistent relations among R_x values for United States whites, 1890 to 1950, for Canada, England and Wales, and the Union of South Africa R_{50} with no exceptions is the largest ratio, and either R_{40} or R_{55} is the lowest ratio with the exception only of England and Wales.

TABLE 9

R_x FOR VARIOUS AGES, MALE AND FEMALE, IN VARIOUS COUNTRIES

Country and decade	Males				Females			
	40	45	50	55	40	45	50	55
U. S. A.—Whites								
1880–1890	882	1 037	1 084	.958	913	1 065	1.084	.941
1890–1900	917	1 003	1 125	917	.958	1.011	1 110	.914
1900–1910	927	1.004	1 104	967	.946	1.036	1.063	.971
1910–1920	931	1 020	1 072	.975	951	1 035	1.065	.978
1920–1930	883	1 039	1 079	981	947	1 025	1 066	.980
1930–1940	.959	1 001	1 090	987	.964	1.020	1.098	.967
1940–1950	972	1 005	1 073	.978	.970	1.005	1 100	.956
U. S. A.—Nonwhites								
1880–1890	761	1 074	1 236	738	.799	1 217	1.367	.775
1890–1900	.726	1 130	1 168	828	769	1 269	1 233	.884
1900–1910	.748	1 067	1 196	.808	.825	1.162	1.224	.856
1910–1920	714	1 064	1 192	.841	768	1.220	1.215	.923
1920–1930	584	1 130	1 186	.821	.783	1.161	1.241	.862
1930–1940	821	.915	1 643	827	.860	1 069	1 816	.753
1940–1950	.921	970	1.436	864	906	1.047	1.663	.779
Canada								
1931–1941	974	1 008	1 092	.985	977	1.016	1.078	1.000
England and Wales								
1921–1931	970	1 005	1 031	1 007	.976	1 016	1 051	1.018
Union of South Africa								
1921–1931	952	1 045	1.053	.958	.958	1 025	1.100	.941
Norway								
1920–1930	994	1 008	1 005	1 004	.995	.998	1.007	1.009
Sweden								
1900–1910	1 003	999	1 000	1 005	997	1.002	.996	1 003
1910–1920	998	999	1 005	1 008	1 004	.996	1 002	1.008
1920–1930	1 005	.999	1 003	1 007	1 001	1 001	1.001	1.001
1930–1940	997	999	1 002	998	1.000	.998	1.004	.995
Switzerland								
1920–1930	.999	.999	1 006	999	1 000	1.002	1.002	1.000

Sources

Population data for the United States U. S Bureau of the Census, U. S. Census of Population, 1950. Vol II *Characteristics of the Population*, Table 39

Life table data for the United States from official life tables for the United States, for 1901, 1910, 1920–29, 1930–39, and 1945

Population data for foreign countries taken either from census publications (for the Union of South Africa, England and Wales, and Canada) or from statistical yearbooks (for Norway, Sweden, and Switzerland) of the country in question.

Life table data for foreign countries taken from United Nations, *Demographic Yearbook, 1948*, Tables 33 and 34.

$R_x = \dfrac{({}_5S_{x+5})_c}{({}_5S_x)_c} \bigg/ \dfrac{({}_5S_{x+5})_m}{({}_5S_x)_m}$, where ${}_5S_x$ represents the fraction of a cohort aged x to $x+5$ at the beginning of a decade that survives $({}_5S_x)_c$ represents the fraction surviving according to census data, $({}_5S_x)_m$ the fraction surviving according to life table data.

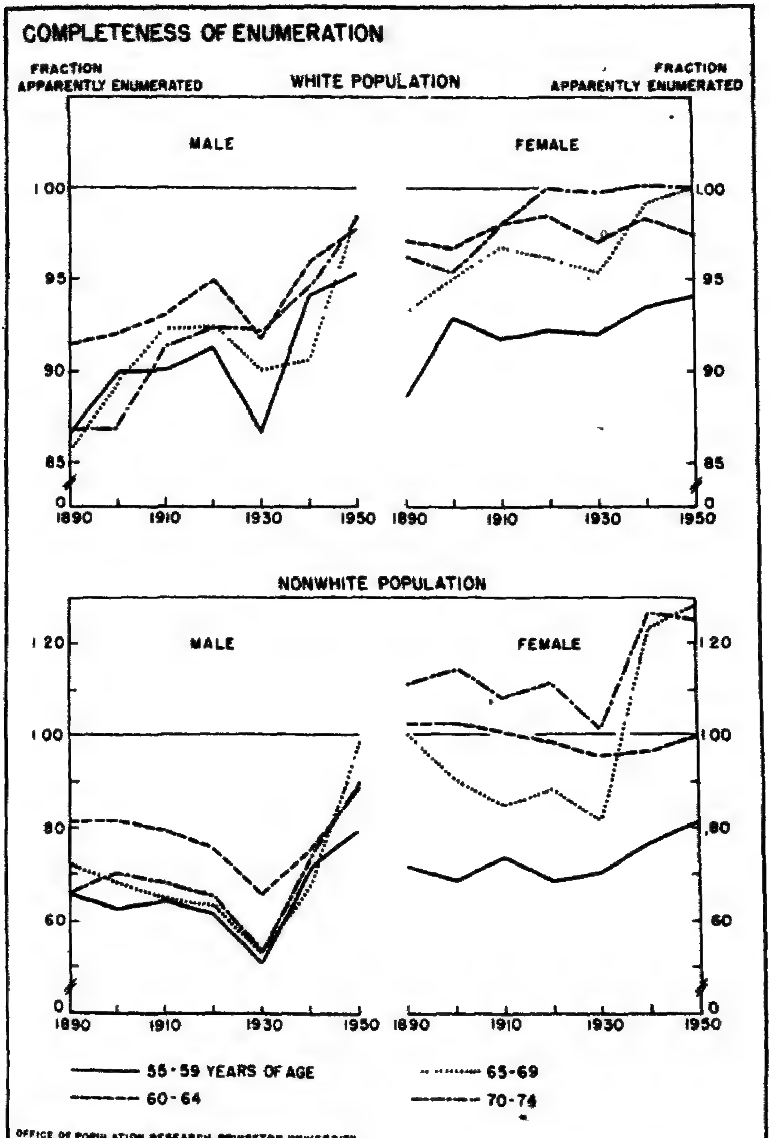


FIG. 7. Apparent completeness of enumeration of persons between the ages of 55 and 75 in U. S. censuses from 1890 to 1950 (based on a comparison of census and life table survivorship ratios) by sex and color.

The R_x values do not as they appear in Table 9 reveal the time pattern of erroneous enumeration in the United States censuses However, if we assume uniform enumeration up to age 50–54 in all censuses, and assume that life table values yield an accurate estimate of ratios of cohort survivorships, it becomes possible to calculate the apparent completeness of enumeration of white males and white females at ages 55–59 through 70–74 for the censuses from 1900 to the present.

The results of such calculations are presented in Figure 7. The *level* of underenumeration in Figure 7 cannot be taken seriously (being based on an assumption for white males of no error in enumeration in all censuses at ages 40–44 and 50–54, and of about 2 per cent underenumeration at ages 45–49), but *changes* probably are significant The fact that 1930 is shown to be more underenumerated than 1940 and 1950 for all of the ages considered—and especially so for the males—is striking

REFERENCES

[1] Eckler, A Ross, and Pritzker, Leon, "Measuring the accuracy of enumerative surveys," a paper presented by A. Ross Eckler at the International Statistical Institute in New Delhi, India, December 5, 1951. U S. Bureau of the Census, processed.

[2] Grove, Robert D , "Studies in the completeness of birth registration . . ." U. S Bureau of the Census, *Vital Statistics—Special Reports*, Vol. 17 (1943)

[3] Metropolitan Life Insurance Company, *Statistical Bulletin*, Vol. 32, No. 8 (1951).

[4] Myers, Robert J , "Errors and bias in the reporting of ages in census data," *Transactions of the Actuarial Society of America*, Vol 41 (Part 2), (1940).

[5] Shapiro, Sam, and Schachter, Joseph, "Birth registration completeness: United States, 1950," *Public Health Reports*, Vol 67 (1952).

[6] Shapiro, Sam, and Schachter, Joseph, "Methodology and summary results of the 1950 birth registration test in the United States," *Estadistica*, Vol. 10 (1952).

[7] United Nations, *Population Bulletin*, No 2 (October 1952), pp. 59–79.

[8] U. S. Department of Commerce, Bureau of the Census, *Historical Abstract of the United States, 1789–1945*, Washington, U. S Government Printing Office, 1949

[9] U. S Department of Commerce, Bureau of the Census, 1950 Census of Population, Vol. II *Characteristics of the Population*, U. S Government Printing Office, 1953.

[10] U S National Office of Vital Statistics, "Abridged life tables, United States, 1950," *Vital Statistics—Special Reports*, Vol 37 (1953).

[11] U. S National Office of Vital Statistics, "Summary of natality statistics, United States, 1950," *Vital Statistics—Special Reports*, Vol. 37 (1953).

[12] U S. National Office of Vital Statistics, *Vital Statistics of the United States, 1949, Part I*, Washington, U. S Government Printing Office, 1951.

[13] U. S. National Office of Vital Statistics, *Vital Statistics—Special Reports*, Vol. 33, No. 8 (1950).

ELEMENTS OF SYMMETRY IN THE SKEWED INCOME CURVE*

HERMAN P. MILLER
Bureau of the Census

The thesis of this paper is that the skewness of income distributions is largely due to merging several symmetrical distributions which differ primarily with respect to level and dispersion Much of the skewness of the income distribution is due to the inclusion of women. Even for men, there is considerable symmetry in income distributions when occupational groups are considered individually. The "tail" of the income distribution largely includes men employed as independent professionals, businessmen, or managers To the extent that there is freedom of entry into these occupations, income differences between these groups and others may merely represent the payment by society for rare skills or services The facts regarding freedom of entry are not now adequately known

NUMEROUS attempts have been made to construct theoretical systems which would account for the skewness of the income curve. These theories can be traced as far back as the latter part of the nineteenth century, and they persist with full vigor to the present day. In general, theories of the distribution of personal income can be classified as either "natural" or "institutional." The natural theories generally take a mathematical form and they attempt to explain income distribution in terms of models which are generated by the theory of probability. Although some of these theories take institutional factors into account, the great majority regard income distributions as the result obtained by a play of natural forces outside the economic system; forces which would persist regardless of the particular form of economic organization These theories range from a view of the income curve as a joint probability distribution of biological traits transmitted by heredity,[1] to the analysis of income distribution in terms of a game of chance played under certain specified conditions.[2]

In contrast to the natural theories, the institutionalist approach tends to be nonmathematical in nature The theories based on this ap-

* This article is an adaptation of part of a chapter from the 1950 Census monograph *Income of the American People*, to be published soon by John Wiley and Sons, Inc This monograph was prepared under the joint sponsorship of the Social Science Research Council and the U. S. Bureau of the Census.

[1] Carlos C. Closson, "Some social applications of the doctrine of probability," *Journal of Political Economy*, 7 (March 1899)

[2] Maurice Fréchet, "Nouveaux essais d'explication de la répartition des revenus," *Revue de L'Institut International de Statistique*, 13 (1945).

proach typically do not seek a model which can produce the skewed income curve. Rather, they try by means of descriptive analysis to explain income inequality in terms of the institutional setting. These theories recognize the importance of certain natural phenomena like the distribution of ability, chance, etc. However, their unique stamp is the emphasis they give to institutional arrangements like the inheritance laws which they claim help to perpetuate inequality once it is established. Thus, for example, Taussig typifies this view when he attributes the origin of inequality to differences in innate ability and the perpetuation of inequality to "the influence of the inheritance both of property and of opportunity," as well as to the biological transmission of "native ability."[3] Pigou states the issues even more succinctly. In his *Economics of Welfare*, he asks why the curve of income is skewed when "there is clear evidence that the physical characters of human beings—and considerable evidence that their mental characters—are distributed on an altogether different plan," i e , a normal curve.[4] Pigou first considers the possibility that the skewed income curve largely reflects the merging of nonhomogeneous groups, each of which can be characterized by a non-skewed curve. However, he ultimately discovers "a more important and more certain explanation"; namely, "Income depends not on capacity alone, whether manual or mental, but on a combination of capacity and inherited property. Inherited property is not distributed according to capacity, but is concentrated upon a small number of persons."[5] Pigou thus attributes the major explanation for the skewness of the income curve to institutional factors which bestow an undue advantage upon the wealthier classes in society.

The evidence presented below for the United States suggests that the skewness of the income curve reflects the interplay of many forces and that it would be unwise and perhaps incorrect to overstress the importance of any single factor. This is not to say that the statistical evidence either supports or contradicts Pigou's conclusions Like all statistics, these data are subject to interpretation and they can be read in such a way as to find them consistent with either the natural or the institutional theories Perhaps the chief value of the empirical evidence is that it focuses attention on the characteristics associated with the component parts of the income curve and permits the formulation of more specific hypotheses regarding the controlling factors in the determination of personal income distribution.

---

[3] F. W. Taussig, *Principles of Economics* (New York Macmillan Company, 1920) Second Edition, p 246.

[4] A. C. Pigou, *Economics of Welfare* (London Macmillan and Co.), Third Edition, 1929, p. 648.

[5] *Ibid.*, p. 649.

TABLE 1

DISTRIBUTION OF THE TOTAL CIVILIAN NONINSTITUTIONAL POPULATION OF THE UNITED STATES BY TOTAL MONEY INCOME: 1951

Total money income	Number of persons (thousands)	Per cent distribution
Total	151,532	100.0
Loss	287	0.2
No income[1]	79,786	52.7
$1 to $499	11,553	7.6
$500 to $999	8,968	5 9
$1,000 to $1,499	5,955	3 9
$1,500 to $1,999	6,314	4.2
$2,000 to $2,499	7,246	4.8
$2,500 to $2,999	6,385	4.2
$3,000 to $3,499	6,959	4 6
$3,500 to $3,999	5,309	3.5
$4,000 to $4,499	3,946	2 6
$4,500 to $4,999	2,296	1 5
$5,000 to $5,999	3,013	2 0
$6,000 to $6,999	1,363	0.9
$7,000 to $9,999	1,220	0.8
$10,000 to $14,999	502	0.3
$15,000 and over	430	0 3

[1] Includes the following groups.

Children under 14 years old	42,204
Persons aged 14–19, most of whom were attending high school or college	7,702
Housewives	25,609
Persons aged 65 and over (excluding housewives)	1,077
Other adults	3,194

Source: Estimated population based on unpublished data of the Census Bureau. Income distribution derived from the Census Bureau report, *Current Population Reports—Consumer Income*, Series P-60, No. 11, table 1.

COMPONENT PARTS OF THE INCOME CURVE

In April 1952 there were about 151½ million civilians residing in the United States. An examination of the distribution of these people by their own incomes indicates that the income curve is very much as Pigou described it, "humped and lopsided." The very striking fact about Figure 1 is that about one-half of the people in the United States received no cash income at all during the calendar year 1951 If persons who did unpaid work on the family farm or business were included as

income recipients the proportion without income would be reduced only slightly since the maximum number of people who were so employed during any given month in that year was less than 2½ million. The proportion without income may be somewhat overstated because of the difficulty of correctly allocating the income within a household and because of the tendency for people to forget to report small amounts of

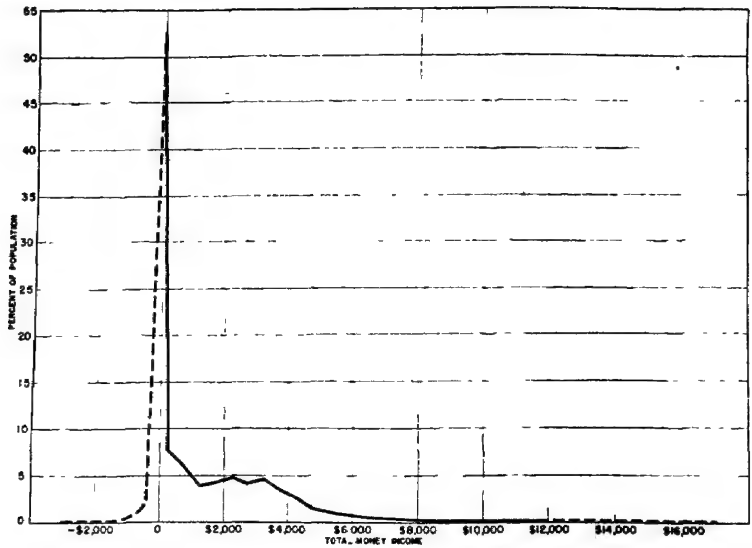


FIG 1 Per cent distribution of the civilian noninstitutional population by total money income, for the United States: 1951.

interest, dividends, and similar types of income to interviewers However, even if all of the imperfections of measurement could be eliminated, it is doubtful that the conclusion that only one-half of the population in this country receives money income would have to be changed significantly.

There are some who will object to the presentation of the statistics as they are shown in Figure 1 They will correctly point out that the 80 million people without income include 42 million children under 14 years of age and 8 million additional persons from 14 to 19 years old, most of whom are going to school. The figure also includes 26 million housewives and 1 million aged persons. Surely these groups should be treated separately if the aim is to understand the forces which shape

TABLE 2

PER CENT DISTRIBUTION OF CIVILIAN NONINSTITUTIONAL INCOME RECIPIENTS 14 YEARS OF AGE AND OVER BY TOTAL MONEY INCOME, BY SEX, FOR THE UNITED STATES. 1951[1]

Total money income	Both sexes	Male	Female
Number of persons with income (thousands)	71,746	46,572	25,174
Per cent	100.0	100 0	100.0
Loss	0.4	0.4	0 2
$1 to $499	16.1	8.9	29 2
$500 to $999	12 5	8.5	19.6
$1,000 to $1,499	8 3	6 8	10 9
$1,500 to $1,999	8.8	6 8	12.3
$2,000 to $2,499	10 1	9 6	11.0
$2,500 to $2,999	8 9	9 8	7.2
$3,000 to $3,499	9 7	12.4	4 9
$3,500 to $3,999	7.4	10 3	2.2
$4,000 to $4,499	5.5	8 1	0.9
$4,500 to $4,999	3 2	4 8	0.5
$5,000 to $5,999	4 2	6 3	0 5
$6,000 to $6,999	1 9	2 8	0.2
$7,000 to $9,999	1 7	2.5	0 2
$10,000 to $14,999	0 7	1 1	0 1
$15,000 and over	0 6	0 9	0.1

[1] To facilitate graphic presentation, all per cents shown in Figure 3 are based on the total of both sexes (71,746) rather than on the individual column totals which were used in this table

Source Derived from U S. Bureau of the Census, *Current Population Reports—Consumer Income*, Series P-60, No. 11, Table 4

income distribution There is considerable dispute as to who should be included in a measure of income distribution. It has been argued that a measure of income inequality should refer to the entire population rather than to a segment of it [6] The major reason for including all of these groups in the first approximation of the income curve is to provide an inventory of the entire population with respect to income.

[6] "In brief, no group less extensive than the total population should be the base for studying changes in the equality of incomes Morris A Copeland came to this conclusion in *Recent Economic Changes in the United States*, and Kuznets has stressed its logic in 'National Income,' *Encyclopedia of Social Sciences*, and *National Income A Summary of Findings* " The above quotation appears in an article by Dorothy S Brady, "Research in the Size Distribution of Income," *Studies in Income and Wealth* (New York. National Bureau of Economic Research) Volume 13, p 10.

Even if the analysis of income distribution is restricted to income recipients, it is apparent from Figure 2 that the income curve remains much the way Pigou described it. Over one-fourth of the income recipients are clustered in the relatively narrow range of incomes between $1 and $1,000. If this range is doubled, nearly half of all income recipients can be accounted for.

It can be noted in Figure 2 as well as in the other figures that the curve for both extremes of the distribution has been fitted with a dotted

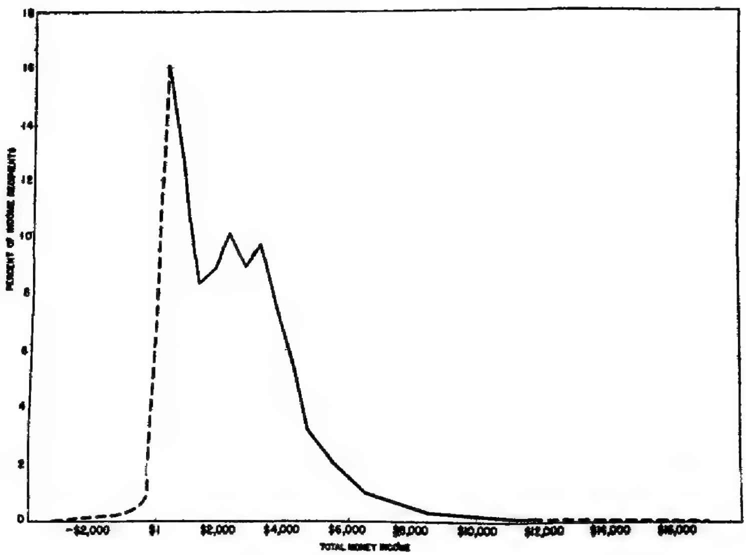


FIG 2. Per cent distribution of income recipients by total money income, for the United States: 1951

line rather than with the solid one which has been used for the rest of the distribution The dotted line has been used to indicate that the data for this part of the curve represent only rough approximations. For example, in Figure 2, the only information available for the "Loss" interval and the "$15,000 and over" interval is that they respectively include 0 4 per cent and 0 6 per cent of the income recipients Since these intervals comprise only a small area of most of the distributions for which data are presented, this liberty with the figures has the advantage of indicating the shape of the complete curve without significantly affecting any of the conclusions.

One of the most important sources of heterogeneity in the income curve is the inclusion of both men and women in the same distribution. It is desirable for several reasons to show income distributions sepa-

TABLE 3

PER CENT DISTRIBUTION OF FEMALE INCOME RECIPIENTS 14 YEARS OF AGE AND OVER BY TOTAL MONEY INCOME AND EXTENT OF EMPLOYMENT, FOR THE UNITED STATES: 1949[1]

Total money income	Worked 50 weeks or more in 1949	Worked less than 50 weeks or did not work at all in 1949		
		Total	Worked 1 to 49 weeks	Did not work
Number of women with income (thousands)	7,981	14,291	9,348	4,943
Per cent	100.0	100.0	100 0	100.0
$1 to $499 or loss	6.0	43.8	42.0	47.9
$500 to $999	9.3	24.0	21 9	27.8
$1,000 to $1,499	14.5	12 5	14.2	9.2
$1,500 to $1,999	21.1	7.6	9 2	4.5
$2,000 to $2,499	22.1	5.0	5.9	3.2
$2,500 to $2,999	12.7	2 4	2 9	1.5
$3,000 to $3,499	7.1	1 6	1.7	1.4
$3,500 to $3,999	2.8	0 8	0.8	0.7
$4,000 to $4,499	1 5	0.6	0.5	0.7
$4,500 to $4,999	0 8	0 3	0.3	0.3
$5,000 to $5,999	0.9	0 5	0.3	0 8
$6,000 to $6,999	0.4	0.2	0.1	0.4
$7,000 to $9,999	0.4	0.3	0.1	0.6
$10,000 and over	0.4	0.4	0.1	1.0

[1] To facilitate graphic presentation, all per cents shown in Figure 4 are based on all female income recipients (22,272) rather than on the individual column totals which were used in this table.

Source: U. S. Bureau of the Census, *U. S. Census of Population. 1950*, Vol. II, *Characteristics of the Population*, Part 1, United States, Table 141.

rately for each sex. Although income from employment represents the most important source of receipts for both men and women, the labor force behavior of women is markedly different from that of men. In our society it is customary for the man to provide for his own support and for that of his family. Accordingly, practically all able-bodied men in

the productive age groups become full-time workers and develop permanent attachments to the labor force. In contrast, it is customary for women to have the primary responsibility for home management. During any given month, roughly three-fourths of the married women do not engage in any paid economic activity and only one-fourth are in the labor force as either paid workers or unpaid workers in their family

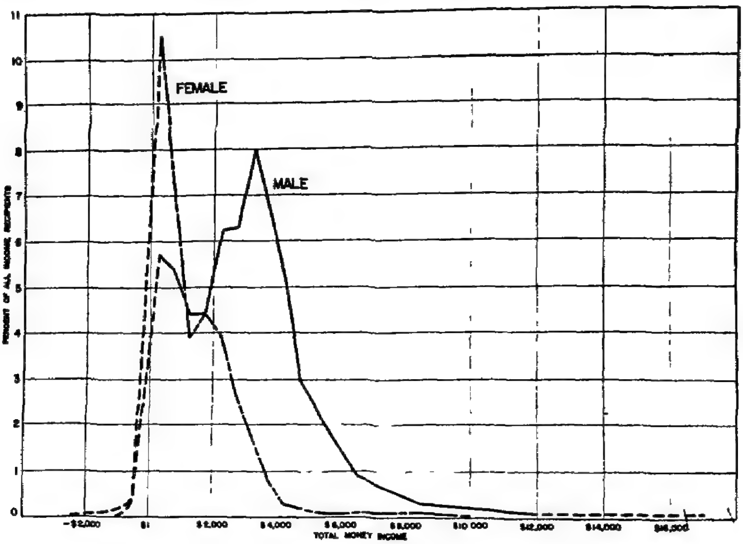


FIG 3 Per cent distribution of income recipients by total money income, by sex, for the United States 1951

farm or business [7] Nevertheless, married women comprise about one-half of the female labor force.[8] This means among other things that the majority of women workers can typically accept only intermittent employment and at jobs which interfere least with the fulfillment of their prime responsibility. Moreover, even when the working woman is not a housewife, she frequently limits herself to certain types of jobs. For these and many other reasons, it is important in the analysis of income distribution that the data for men and women be studied separately.

Figure 3 presents the separate income distributions for men and

[7] U S Bureau of the Census, *Current Population Reports—Labor Force*, Series P-50, No. 39 Table 1

[8] *Ibid.*, Table 1

women. It will be noted that both curves are skewed and show a pronounced tendency toward bimodality. Neither curve resembles the "symmetrical curve shaped like a cocked hat" which Pigou sought. In terms of the problem which Pigou set out to analyze, it should be noted that the skewness of the income curve for women can largely be explained in terms of the factors noted above rather than by reference to inheritance. In interpreting Figure 3 and the subsequent figures which will be discussed, it should be noted that the combined area under both curves represents unity. The area under the curve for males represents nearly two-thirds of the total area under both curves because nearly two-thirds of all income recipients were men These curves as shown, therefore, accurately reflect the weights attached to each of the components of the distribution

One important fact to note about female income recipients is that when the distributions for full-year workers are separated from the distributions for women without work experience during the year or for women who worked only part of the year, two unimodal curves appear. The arithmetic mean and the median of the curve for the full-year workers differ by only $100[9] and an examination of the relationship between the quartiles and median, which is a rough measure of the symmetry of a distribution, indicates only moderate skewness [10] The curve for the nonworkers and the part-year workers is considerably more skewed than the curve for the full-year workers The quartiles in this distribution are not symmetrical about the median;[11] and the mean is about $400 higher than the median indicating a concentration of people in the lower part of the distribution [12] This distribution resembles the type of income curve which Pigou had in mind. Much of the asymmetry of the curve for nonworkers and part-year workers can be explained in terms of the combination of dissimilar groups such as nonworkers, many of whom were living on transfer payments or on inherited property, and part-year workers whose periods of employment could have ranged anywhere from 1 week to 49 weeks It is equally important to note, however, that the small absolute difference between the mean and the median indicates that the tail of the distribution carries relatively little weight Therefore, from an analytical viewpoint, the essential feature of this distribution may well be the symmetry which characterizes it throughout most of its range.

---

[9] Arithmetic mean, $2,109; Median, $1,980

[10] $1-(Q_1/Q_2) = .33$; $(Q_3/Q_2) - 1 = .30$.

[11] $1-(Q_1/Q_2) = .55$; $(Q_3/Q_2) - 1 = 1.05$.

[12] Arithmetic mean, $1,025; Median, $628.

The income curve for men in Figure 3 retains some of the skewness and much of the bimodality which appeared in the distribution for all income recipients Apparently, therefore, this curve may still contain elements of heterogeneity which are to be accounted for. The simple expedient of classifying income recipients as either full-year workers or as nonworkers or part-year workers, which was used for women, cannot

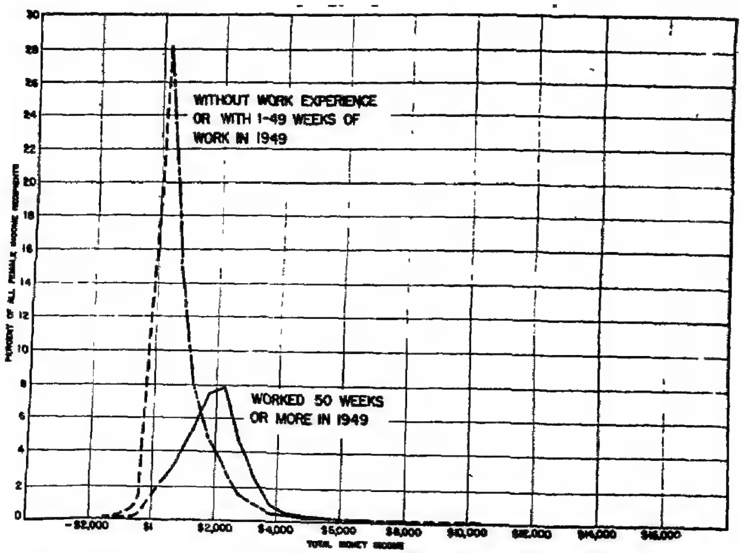


FIG. 4. Per cent distribution of female income recipients by total money income and extent of employment, for the United States: 1949.

be meaningfully employed for men. Although the nonworkers do represent a significantly different group from full-year workers, the same meaning can not necessarily be ascribed to the part-year workers. A large proportion of the men who worked less than a full year may have either worked in industries like construction where pay scales are adjusted to account for the seasonality of the work, or else they experienced some temporary unemployment. As a first approximation, it is convenient to regard the income curve for men as a combination of the curves for two groups: (*a*) those who are employed; and (*b*) those who are not employed, which includes both those who are unemployed and those who are not in the labor force. Although separate data are available for the unemployed, they are combined with persons who were not

in the labor force because of their relatively small numbers. The inclusion of the unemployed with the employed would not have significantly changed the results.

TABLE 4

PER CENT DISTRIBUTION OF CIVILIAN MALE INCOME RECIPIENTS 14 YEARS OF AGE AND OVER BY TOTAL MONEY INCOME IN 1951, BY LABOR FORCE STATUS IN APRIL 1952, FOR THE UNITED STATES[1]

Total money income	Employed	Unemployed or not in the labor force		
		Total	Unemployed	Not in the labor force
Number of men with income (thousands)	40,687	5,885	950	4,935
Per cent	100 0	100.0	100 0	100.0
Loss	0.5	0.2	0.4	0.2
$1 to $499	5.1	33.7	21.7	36 0
$500 to $999	5.3	29.3	16.3	31 8
$1,000 to $1,499	6.0	12.6	14.1	12.3
$1,500 to $1,999	6.6	8.0	14.9	6.7
$2,000 to $2,499	10.4	4.7	9.4	3 9
$2,500 to $2,999	10.8	3.4	8.7	2.3
$3,000 to $3,499	13 9	2.5	5.4	1.9
$3,500 to $3,999	11.7	1.3	1 1	1.4
$4,000 to $4,499	9.1	1.1	4.0	0.6
$4,500 to $4,999	5.4	0.6	2 2	0.3
$5,000 to $5,999	7.1	0.9	0.7	0.9
$6,000 to $6,999	3.2	0 5	0.7	0 4
$7,000 to $9,999	2.8	0.8	0.4	0 8
$10,000 to $14,999	1.2	0 3	—	0.4
$15,000 and over	1.0	0 1	—	0.1

[1] To facilitate graphic presentation, all per cents shown in Figure 5 are based on all male income recipients (46,572) rather than on the individual column totals which were used in this table.

Source: U. S. Bureau of the Census, *Current Population Reports—Consumer Income*, Series P-60, No. 11, Table 4.

It is apparent from Figure 5 that this dichotomy eliminates much of the bimodality from the income distribution for men. The income curve for men who were not employed in April 1952 is quite symmetrical.

Although the curve for employed men is more symmetrical than the curve for all men shown in Figure 3, it apparently still contains some elements of heterogeneity. Note, for example, the pronounced bulge in the lower part of the distribution as well as the extended tail of the distribution in the direction of the higher values. In order to obtain a better understanding of the income distribution for employed men, it is necessary first to examine the distributions for the component occupation groups In Figure 6 the income distribution for employed men has

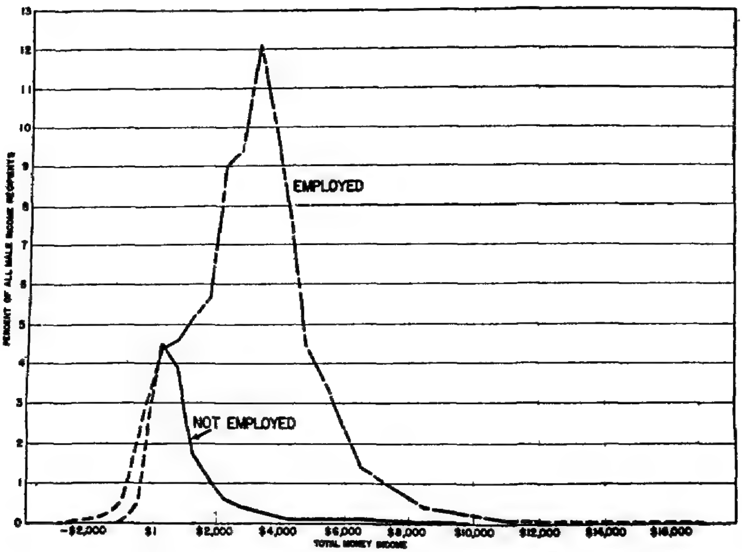


FIG 5 Per cent distribution of male income recipients by total money income in 1951, by employment status in April 1952, for the United States

been divided into three groups: (a) farmers and farm managers (9 per cent of the total); (b) service workers and laborers (17 per cent of the total); and (c) men employed in other occupations (74 per cent of the total). It is apparent that the distributions for the first two groups are asymmetrical The distribution for farmers and farm managers is skewed in the direction of the higher incomes, and the distribution for service workers and laborers has several modal groups. In contrast, there appears to be considerable symmetry in the income distribution for other employed men despite the fact that it is an amalgamation of many different occupation groups Before proceeding with an analysis

of the occupation groups included in the "other employed males" category, it is important to point out that further analysis may provide an explanation for the asymmetry of the distributions for farmers and for

TABLE 5

DISTRIBUTION OF EMPLOYED MALES 14 YEARS OF AGE AND OVER BY TOTAL MONEY INCOME IN 1951, BY OCCUPATION GROUP IN APRIL 1952, FOR THE UNITED STATES[1]

Total money income	Service workers and laborers	Farmers and farm managers	Other employed men			
			Total	"White collar" workers[2]	"Blue collar" workers[3]	Independent professionals, nonfarm proprietors, and managers and officials
Number of men with income (thousands)	6,915	3,795	29,977	7,540	16,962	5,475
Per cent	100.0	100.0	100.0	100.0	100.0	100.0
Loss	0.2	2.6	0.3	0.1	0.1	1.3
$1 to $499	10.6	16.5	2.4	3.7	1.8	2.6
$500 to $999	10.0	15.7	2.9	3.3	2.9	2.2
$1,000 to $1,499	11.6	14.9	3.6	3.0	4.0	3.2
$1,500 to $1,999	12.5	11.1	4.7	3.6	5.7	3.0
$2,000 to $2,499	15.1	10.4	9.3	6.6	11.0	7.2
$2,500 to $2,999	11.7	6.7	11.1	9.6	12.6	8.1
$3,000 to $3,499	13.3	5.3	15.2	14.4	16.8	10.1
$3,500 to $3,999	7.2	2.9	13.8	14.7	14.7	9.4
$4,000 to $4,499	3.6	2.9	11.1	12.5	11.3	8.7
$4,500 to $4,999	2.3	1.3	6.6	7.0	6.5	6.3
$5,000 to $5,999	1.1	3.0	9.0	9.3	8.3	11.3
$6,000 to $6,999	0.2	1.6	4.1	5.5	3.0	5.8
$7,000 to $9,999	0.4	2.2	3.4	5.0	1.1	9.1
$10,000 to $14,999	—	1.2	1.4	1.1	0.2	6.2
$15,000 and over	0.2	1.9	1.1	0.6	—	5.5

[1] To facilitate graphic presentation, all per cents shown in Figure 6 are based on all employed males with income (40,687) and those in Figure 7 are based on the total "other" employed males (29,977) rather than on the individual column totals which were used in this table.

[2] Salaried professional and technical workers, clerical workers, and sales workers.

[3] Craftsmen, foremen, operatives and kindred workers.

Source U S Bureau of the Census, *Current Population Reports—Consumer Income*, Series P-60, No. 11, Table 5

service workers and laborers. In the case of the farmers, for example, it is entirely possible that the single skewed distribution could be reduced to two symmetrical distributions if separate data were available for sharecroppers and for other farmers. In the case of service workers

and laborers preliminary analysis indicates that even the more detailed occupations within these groups such as farm laborers, nonfarm laborers, and service workers have asymmetrical income distributions. It may well be that variations in extent of employment are a key factor in the explanation of the asymmetry of these distributions.

Returning once again to the over-all income distribution for employed men in Figure 5, it is apparent that the bulge in the lower part of the distribution is largely attributable to the inclusion of farmers,

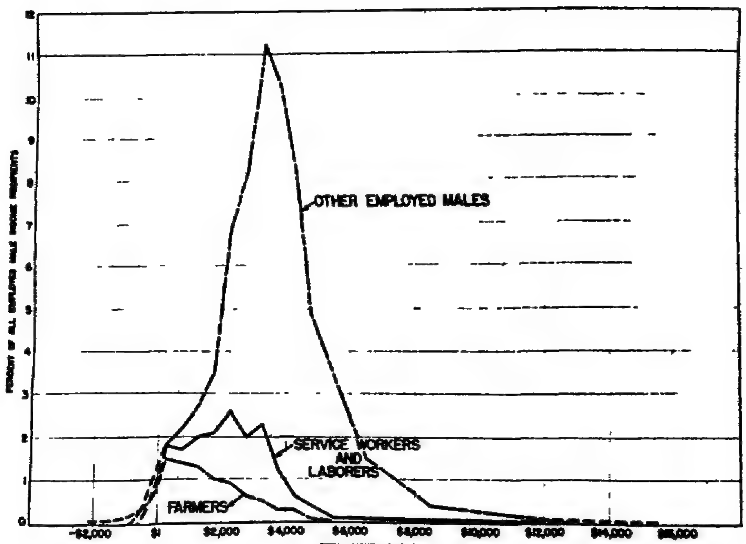


FIG. 6. Per cent distribution of employed males by total money income in 1951, by occupation group in April 1952, for the United States.

service workers, and laborers Once these occupation groups are removed for separate analysis, the distribution for the remaining three-fourths of the employed men is quite symmetrical except for the fact that it retains a rather pronounced "tail" in the direction of the higher values. In order to achieve a better understanding of this distribution, it has been divided into three occupation groups: (*a*) "blue-collar" workers or operatives and craftsmen (57 per cent of the total); (*b*) "white-collar" workers or salesmen, clerks, and salaried professionals (25 per cent of the total); and (*c*) independent professionals, nonfarm proprietors, and managerial workers (18 per cent of the total). Here it

may be noted that the three distributions are essentially symmetrical and appear to differ from each other primarily with respect to level of income and income dispersion. The "tail" of the over-all distribution is largely attributable to the independent professional, business, and managerial group which contains about three-fourths of all men with incomes over $10,000.

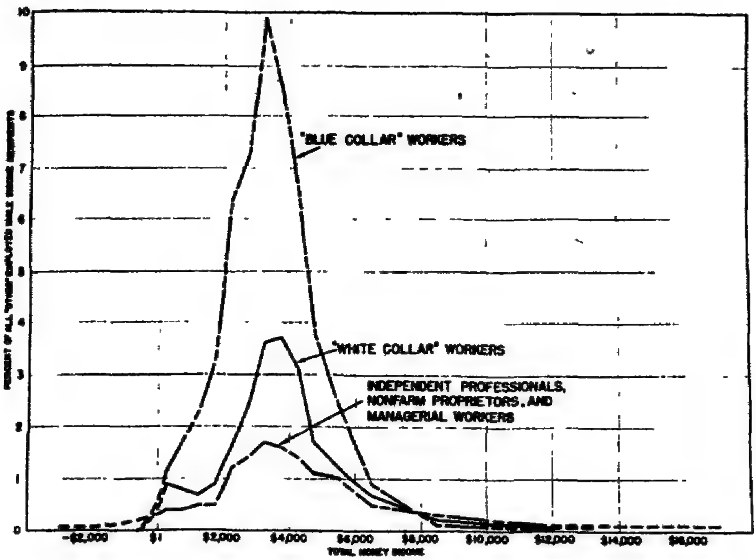


FIG. 7. Per cent distribution of "other" employed males by total money income in 1951, by occupation group in April 1952, for the United States.

SOME TENTATIVE CONCLUSIONS

In this paper an attempt has been made to explain the skewness of the over-all distribution of income in terms of the major components of that distribution rather than by reference to special institutional conditions such as the advantages bestowed by inheritance. The basic thesis is that much of the skewness of the income curve is primarily attributable to the merging of several different types of distributions which are themselves largely symmetrical. It has been shown that much of the skewness of the income curve is attributable to the inclusion of women in the distribution. Although the essential difference between the income curves for men and women may stem from the mores of our society, it has little to do with inheritance. Even in the

case of men, there is considerable symmetry in income distribution. The income distribution for the three groups of occupations which together include about three-fourths of the employed men were found to be quite symmetrical when analyzed separately

Once the factors which largely account for the skewness of the income distribution in the United States are separated, it immediately becomes apparent that the skewness of income distribution for other times and places may have resulted from an entire different set of factors In less industrialized societies the landed aristocracy would undoubtedly replace the professional and managerial groups as the upper income groups Therefore, two societies may have the same degree of inequality of incomes for different reasons.

The data which have been presented do not in themselves either prove or disprove the conclusion reached by Pigou. Even if the skewness of the income curve can be explained in terms of the merging of nonhomogeneous groups, the fact nevertheless remains that there are great variations in the incomes received for different kinds of work Inheritance may be an important factor in explaining income inequality in a number of obvious ways, including the restriction of entry into the better-paying occupations In 1951 only 2 per cent of the men made over $10,000, but they received 12 per cent of the income [13] It is significant that only a very small proportion of these men can be classified as the "idle rich " For example, 99 out of every 100 of these men were in the labor force in April 1952,[14] and 70 out of every 100 derived their incomes entirely from earnings.[15] Only 2 per cent of these men derived their incomes entirely from property or investments [16] Further analysis of the characteristics of the highest income recipients indicates that nearly three-fourths of them were either independent professionals, businessmen, or managerial workers [17] If farmers are included, about 85 per cent of the highest income group can be accounted for It is apparent, therefore, that these occupations are the channels through

[13] Estimated from Census Bureau report, Series P-60, No. 11, Table 1.

[14] *Ibid*, Table 4

[15] *Ibid*, Table 7 Earnings as defined here include income received from wages or salaries or from the operation of a farm, business, or professional practice It is undoubtedly true that in a broad philosophical sense this definition of earnings may include the receipts from property or investments in which work was performed only in a nominal way. However, it is impossible to determine the magnitude or the importance of such receipts from the data which are currently available.

[16] These figures undoubtedly understate the prevalence of investment income within this group They also probably exaggerate the extent of employment among the wealthy because of the inclusion of some people who hold purely nominal jobs Nevertheless, there is no evidence to controvert the conclusion that earnings are the most important source of income for the upper income groups.

[17] Census Bureau report, Series P-60, No. 11, Table 5.

which one can "hit the jackpot" in our society.[18] If it can be shown that entry into the independent professions, business ownership, or managerial work is limited to persons reared in wealthy families, much of Pigou's analysis could be correct. However, if there is relatively free entry into these occupations, then it might well be that "a large part of the existing inequality of wealth can be regarded as produced by men to satisfy their tastes and preferences."[19] In other words, to the extent that all men have access to the professional and managerial occupations, and the numbers admitted to these occupations are sufficient to keep monopolistic practices at a minimum, the income differences between these occupations and others may merely reflect the payments by society for rare skills or risk-taking.

---

[18] Some will argue that an annual income of $10,000 is a rather small "jackpot." The data in a recent Census Bureau report (Series P-60, No. 14) indicate that essentially the same conclusions apply to the "$25,000 and over" income group.

[19] M. Friedman, "Choice, chance, and the personal distribution of income," *Journal of Political Economy*, LXI (August 1953).

A STRUCTURE OF MONEYFLOWS*

MORRIS MENDELSON
Pennsylvania State University

THERE has been general agreement among the reviewers that *A Study of Moneyflows in the United States*[1] is a significant contribution to economics and deserves careful consideration. This paper is intended to explain the broad outlines of Professor Copeland's moneyflows accounts and discuss the general problems involved. It also makes the basic concepts more accessible to the general reader.

Copeland's accounts differ from both the gross national product accounts of the National Income Division (NID) of the Department of Commerce and the moneyflows accounts that have been subsequently developed from Copeland's study by the research staff of the Board of Governors of the Federal Reserve System. In general the events embraced by each particular set of accounts are delineated by the problems the accounts are to deal with. There can be no direct comparison of the merits of the different social accounting systems and no such comparison will be attempted here.

In most of the events with which we concern ourselves in economic analysis there is a surrender of an economic good (or service) for a financial claim of some sort.[2] The surrender of economic goods may be looked upon as a flow of values and the corresponding financial claims as the synchronous compensatory flow. While, with few exceptions, the first type of flow is always matched by a synchronous compensatory flow, it is not necessarily true that the matching flow is one of financial claims. There are also cases of two way flows of financial claims.

Every flow has an origin and a terminus, i.e., an economic organism from which the flow originates and another at which it terminates. The

* I am indebted to Mr. D. H. Brill and Mr. S. J. Sigel of the Research Staff of the Board of Governors of the Federal Reserve System both for their extensive and constructive criticism and for the many ideas I have absorbed from their unpublished manuscripts which they so kindly made available to me. It proved impossible to give adequate acknowledgment to these papers and the extent to which they are cited below in no way indicates how heavily I have relied upon them. I am also indebted to Professor M. A. Copeland of Cornell University and Mr. J. C. Dawson of the University of Maryland for their extensive comments, criticisms, and suggestions. I alone, however, am responsible for whatever errors remain.

[1] M. A. Copeland, *A Study of Moneyflows in the United States*, National Bureau of Economic Research, New York, 1952. Henceforth this book will be referred to as *Moneyflows*.

[2] For an elaboration of this idea see Chandler Morse, *Basic Concepts of Accounting and Their Economic Application*, Chapter 1, Norton Printing Co., Ithaca, N. Y., 1952. Financial claims are to be understood as including currency and bank deposits. As used here, financial claims include currency and deposits, book credit, securities, mortgages, and monetary metal.

origin of one flow is the terminus of the compensatory flow. The size of the flow at the terminus is equal to the size at the origin. If records were kept of all these flows, we would end up with a fourfold register of every event. Every economic organism would register two flows simultaneously. Social accounting becomes possible when events can be characterized by such synchronous compensatory flows of values. This is the double entry aspect of social accounting. This possibility of such fourfold records of events makes possible the systematic organization of data describing the events.

The different sets of social accounts are summaries of the two way flows of different groupings of the events and as such present different perspectives of the economy. It makes little sense to ask of any set of social accounts why it is superior to any other set unless they are both designed to deal with the same problems It is very easy to forget that social accounting is a means rather than an end in economics. There is no set of accounts that can be held inviolable. The research staff of the Board of Governors of the Federal Reserve System, for instance, has developed a somewhat different structure of the Moneyflows accounts than that originally conceived by Copeland because of the different focus of their analytic interests.

Broad Aspects of Moneyflows

Moneyflows are the flows and synchronous compensatory flows that result from a particular set of events The events dealt with in this paper are those that Copeland believes relevant to the problems he poses. More specifically, among these problems are:

"Who purchases the gross national product?

"Where does the money to buy the gross national product come from?

"What part do cash balances play in the process of business expansion and contraction?

"What is the role of the banking and monetary system?"[1]

Even this small sample of the questions that a structure of moneyflows can help us answer should make it abundantly clear that such a set of accounts is highly useful. This alone more than justifies an examination of the accounts. It is with the accounts that this paper is concerned, not with the answers.

The events that give rise to moneyflows can be described as only those in which one flow results in an increase in some financial claim of one party and a corresponding decrease in those of a distinct second

[1] Copeland, *op. cit.*, p. 5.

party.[4] Monetary gifts will be included in moneyflows though, in fact, it is not always possible to measure this latter flow. There must be two distinct parties Each party is called a transactor and each event is called a transaction

An increase in a transactor's financial claims is called a disposition of money and his corresponding counterflow a source of money. These correspond to debits and credits respectively Similarly a decrease in financial claims owned is called a source of money and the corresponding flow a disposition.

"Moneyflows are sources and dispositions of money."[5] Each transaction is taken in its totality so that in the case of each transactor and transaction, the accounts, recording sources and dispositions of money, balance except for statistical discrepancies and deviations from accounting uniformity In the case of each transactor there are two major sources and dispositions of money· financial and non-financial For example, a household's source may be the sale of a second hand good (non-financial) or the sale of a government bond (financial). Obviously only the former results in net changes in financial claims.

Just as each transaction results in an equal source and disposition of money to a transactor,[6] each flow in the transaction is a source of *money to one transactor and an equal disposition to the other* We can thus draw up several types of balancing accounts which, except for statistical discrepancies and deviations from accounting uniformity, should balance [7]

The main circuit moneyflows is a subgroup of moneyflows transactions Copeland suggests that the transactions to be included in the main circuit moneyflows are those that play a substantive part in affecting over-all economic adjustments [8] This criterion must be a loose one, and Copeland does not consider it the only one.

The working definition of the main moneyflows circuit is operational It is the statement of detailed specifications of the measurement of the variables. Among those transactions omitted are what Copeland calls

---

[4] It should be noted that transactions in which the compensatory flows have the opposite effect on financial claims (so that the net effect is zero) are moneyflows I have gathered from discussion that this description of moneyflows, while technically correct violates the spirit of moneyflows. It seems that barter transactions undertaken in the spirit that the financial liabilities resulting from the transactions are cancelled against each other, should be included in moneyflows measures though the magnitude of such transactions may sometimes be difficult to measure.

[5] Copeland, *op cit* p 232

[6] *Ibid.*, p. 234 See implications 2 and 3

[7] *Ibid*, p 232 See *First Feature*

[8] *Ibid.*, p. 10.

(*a*) money changer transactions, (*b*) agency transactions, and (*c*) financial turnover transactions. As we can see from transaction number 14, Table I, all financial transactions are not completely excluded from the system of accounts The net effect of financial transactions are recorded in the loanfund balance accounts, i e., the statements of financial assets and liabilities.

Although moneyflows do not have to be presented in the form of social accounts, they can be, and Copeland has presented them that way. To construct such a system using actual data, as Copeland did, necessitates explicit definitions of the sectors and the transaction types There is, however, nothing final about either the definitions or the arrangement of the data These can and should be changed to reflect the institutional complex and other factors appropriate to the problems at hand.

In this paper, while the differences between the two sets of moneyflows accounts will sometimes be noted, we shall pay more attention to the differences and similarities between Copeland's accounts and the gross national product accounts

However before embarking upon any extensive detail of the moneyflows structure, we may cursorily examine how even a very condensed version gives us an insight into business fluctuations. For this purpose we may leave Professor Copeland's sector untouched However for the particular illustrative analysis which we wish to undertake here we do not need the detailed transactions, and we shall condense them considerably. Of the fourteen types of transactions (the first fourteen row captions of Table I) into which Professor Copeland has classified moneyflows, the first thirteen are collectively labelled ordinary transactions. For our purposes the transactions may be rearranged and condensed as follows:

1. GNP or Final Product Expenditures	Ordinary Expenditures	5. Product Receipts	Ordinary Receipts
2. Non-Final Product Expenditures		6. Transfer Receipts	
3. Transfer Expenditures		7. Money Obtained Thru Financing	
4. Money Advanced			
TOTAL DISPOSITION OF MONEY		TOTAL SOURCES OF MONEY	

TABLE I

STRUCTURE OF THE MONEYFLOWS ACCOUNTS

Sources & Dispositions of Money—1939—Millions of Dollars

Transactions	Sectors								
	I. Households		II. Farms		III Industrial corp.		IV. Business prop et al.		
Non-Financial	S	D	S	D	S	D	S	D	
1. Gross Cash Pay	45,100	860		780		25,200		7,800	
2. Cash Interest	2,700	1,320		540	340	1,800	60	320	
3. Cash Dividends	3,800				960	4,300	60		
4. Net Owner Takeouts	9,300			2,420				6,160	
5. Installments to Contractors		1,040		60	2,200	680	3,300	700	
6. Gross Rents		3,900		440	520	1,860	220	1,500	
7. Customers Moneyflows	140	51,200	7,900	3,500	119,500	81,500	50,400	36,600	1,
8. Net Payments for Real Estate	600		100			100		100	
9. Taxes Collected		3,200		480		6,500		1,900	5,
10. Tax Refunds	20				40		10		
11. Insurance Premiums		4,300		100		1,060		440	
12. Insurance Benefits	3,720		60		260		140		
13. Public Purpose Payments	1,560	1,130	810		40	30	1,000	60	
14. Financial Net Increases (+) or Decreases (−) Currency & Deposits Assets −									
+		2,500		200		1,000		200	
Liabilities +									
−									
Book Credit ... ... A −									
+						1,200		300	
L +	200				1,000		400		
−				100					
National Gold Account A −									
+									
L +									
−									
Federal Obligation Acc't. A −									
+		200							
L +									2,
−									
Treasury Currency Acc't. A −									
+									
L +									
−									
Other Loans, Securities & Debts Payable A −	900								
+						100			
L +	700						700		
−				200		100			
Corporate Paid-In Capital A −									
+									
L +									
−						200			
Valuation Adjustments Valuation Gains	500								
Valuation Losses						600		200	
TOTAL SOURCES	69,200		8,800		124,900		56,200		10,
TOTAL DISPOSITIONS[3]		69,700		8,800		126,100		56,300	
Discrepancies Sources Not Accounted For Dispositions Not Accounted For	400				1,300		200		

[1] Less than .5 million or between ± .5
[2] Only starred discrepancies are computed from the data in this table The remaining discrepancies are taken from the original
[3] Total sources and dispositions may differ because of rounding even after adjusted for sources and dispositions not accoun
[4] Includes withdrawals of interest accrued on dividends left on deposit.
[5] Includes sinking and trust funds.

TABLE I—(*continued*)
STRUCTURE OF THE MONEYFLOWS ACCOUNTS
Sources & Dispositions of Money—1939—Millions of Dollars

Sectors													
ate and cal rnment	VII. Banking		VIII. Life ins.		IX. Other ins		X. Securities & real estate		XI. Rest of world		Total		Discrepanc
D	S	D	S	D	S	D	S	D	S	D	S	D	S
											45,100		
4,160		580		420		240		1,060				45,100	
	1,700		980		220		900		30		7,400		
560		440						1,300		160		7,400	
	20		20		80		1,100		180		6,200		
		240		20		130		1,300		320		6,200	
											9,300		
								700				9,300	
											5,500		
1,500		40		10		10		1,100				5,500	
	100		200		20		6,900				8,000		
10		50		30		30		240				8,000	
	360				10		1,200		3,200		185,200		
2,800		220		380		810		2,400		3,810		185,000	
			−20								680		
								480		10		680	
											13,700		
		160		130		110		1,300				13,700	
											80		
												80	
			3,640[a]		2,600[a]						6,600		
		90		120[4]		70		380				6,600	
	30						140				4,300		
220				2,160		1,200						4,300	
									180		6,800		
3,240		0[1]						0[1]		40		6,800	
				100		100		100		1,200		4,900	
	5,200										5,200		
												1,500	
											1,500		100
		3,100										3,100	
									3,000		3,000		200
		1,400		500		100						2,200	
											2,200		200
		200										200	
													200
							600						
100[a]		600		800						1,000		900	
											1,200		
								800					
							300						
		100											
											500		1,340*
		100		100		40		100				1,840	
	7,400		4,800		3,000		11,200		6,600				660*
12,600		7,300		4,800		2,840		10,700		6,500			
	0[1]										2,000		
400				0[1]		0[1]		500				900	

remium receipts are gross of cash dividends withdrawn by policyholders. Premium payments are shown net of such dividends. Con private insurance companies to pension funds they maintain for their own employees have been deducted.

Based on the thirteen national moneyflows accounts, the eleven statements of payments and balances, and the various lo accounts in *Moneyflows*. These accounts show the sources for every figure used.

TABLE IIA

SYNOPSIS OF GNP EXPENDITURES, LOAN FUND FINANCING AND OTHER MONEY FLOWS

(Billions of Dollars)

		1936	1937	1938	1939	1940	1941	1942	Source	
	Households									
A	GNP Expenditures	55.6	60.0	57.5	60.7	65.1	74.4	80.0	GNP 101 K	A
B	Net Money Advanced or Returned to Others	2.3	.1	−.4	.3	−.1	5.0	19.6	Tbl 32 L or A	B
C	Net Transfer Payments	−1.6	1.5	.5	−.1	³	1.1	4.2	GNP 102 L	C
D	Dispositions of Money	56.3	62.4	57.6	60.8	65.1	80.5	103.8	A+B+C+(G when G>zero)	D
E	Distributive Share Receipts, etc.[1]	56.0	62.4	56.6	60.4	65.0	80.0	100.2	GNP 101 T	E
F	Sources of Money	56.3	62.4	57.6	60.8	65.1	80.5	103.8	E+(−G when G<zero)	F
G	Discrepancy Included Above (Disposition)[2]	−.3	.9	−1.0	−.4	−.1	−.6	−3.6	P&B I, Footnotes 1 and 2	G
	Farms									
H	GNP Expenditures	1.9	1.5	.9	.9	1.1	1.6	2.1	GNP 103 F	H
J	Net Money Advanced or Returned to Others	.3	−.3	−.1	.6	−.2	.4	1.8	Tbl 32 M or B	J
K	Dispositions of Money	2.2	1.2	.8	1.5	.9	2.0	3.9	H+J	K
L	Product Receipts minus Expenditures n e c[1]	1.9	.9	.4	.7	.2	1.5	3.3	GNP 103 S	L
M	Net Transfer Receipts	.2	.3	.4	.8	.7	.5	.6	GNP 103 X	M
N	Sources of Money	2.2	1.2	.8	1.5	.9	2.0	3.9	L+M	N
	Industrial Corporations									
P	GNP Expenditures	5.3	6.3	2.5	4.1	6.9	9.8	4.8	GNP 105 R	P
Q	Net Money Advanced or Returned to Others	.3	−1.1	.2	2.2	1.6	3.6	10.2	Tbl 32 N or C	Q
R	Net Transfer Payments	.8	.8	.8	.8	.9	1.0	1.4	GNP 106 S	R
S	Dispositions of Money	6.5	6.8	4.9	7.1	9.3	4.4	16.3	P+Q+R+(V when V>zero)	S
T	Product Receipts minus Expenditures n e c[1]	6.5	6.8	4.9	5.9	8.5	14.1	15.8	GNP 106 L	T
U	Sources of Money	6.5	6.8	4.0	7.1	9.3	14.4	16.3	T+(−V when V<zero)	U
V	Discrepancy Included Above (Disposition)[2]	.1	.8	1.4	−1.3	−.8	−.4	−.5	P&B III, Footnotes 1 and 2	V
	Business Proprietors and Partnerships *et al*									
W	GNP Expenditures	2.9	2.9	2.1	2.6	3.3	3.7	2.2	GNP 108 Q	W
X	Net Money Advanced or Returned to Others	.7	0³	.8	−.3	.5	.3	4.3	Tbl 32 P or D	X
Y	Dispositions of Money	3.6	3.6	3.3	2.3	3.8	3.9	7.6	W+X+(c when c>zero)	Y
Z	Product Receipts minus Expenditures n e c[1]	2.8	2.8	2.7	1.5	2.7	2.9	6.7	GNP 109 L	Z
a	Net Transfer Receipts	.8	.7	.6	.6	.7	.6	.8	GNP 109 S	a
b	Sources of Money	3.6	3.6	3.3	2.3	3.8	3.9	7.6	Z+a+(−c when c<zero)	b
c	Discrepancy Included Above (Disposition)[2]	−.1	.7	.4	−.2	−.5	−.4	1.1	P&B IV, Footnotes 1 and 2	c
	Federal Government									
d	GNP Expenditures	4.9	4.4	5.1	5.1	5.9	15.6	54.3	GNP 111 Q	d
e	Net Transfer Payments	2.8	−.4	.1	1.1	1.0	0³	−2.8	GNP 112 R	e
f	Dispositions of Money	7.9	4.3	5.3	6.3	6.9	16.1	51.5	d+e+(j when j>zero)	f
g	Product Receipts minus Expenditures n e c[1]	2.8	4.0	4.0	3.9	4.3	6.0	10.1	GNP 111 e	g
h	Net Money Obtained thru Financing	5.0	.3	1.3	2.2	2.4	10.1	41.2	Tbl 32 E	h
i	Sources of Money	7.9	4.3	5.3	6.3	6.9	16.1	51.5	g+h+(−j when j<zero)	i
j	Discrepancy Included Above (Disposition)[2]	.1	.3	.1	−.1	−.2	.5	−.3	P&B V, Footnotes 1 and 2	j

[1] Receipts from all product and service transactions (including business taxes) minus non GNP, nontransfer expenditures

[2] A n[illegible] tive entry is r[illegible] arded [illegible] a source of mo[illegible]y [illegible]

[3] Lies between ± $50 million

Note Due to rounding, figures for various lines calculated by formulas given in the source column may differ slightly fr[illegible]

Reproduced with permission from M A Copeland, *A Study of Moneyflows in the United States*, National Bureau of Economic Research, New York, 1952.

(Billions of Dollars)

		1936	1937	1938	1939	1940	1941	1942	Source	
	State and Local Governments									
A	GNP Expenditures	6.8	7.2	7.6	8.0	8.2	8.1	7.9	GNP 114 I	A
B	Net Money Advanced or Returned to Others	.3	.2	.3	−.4	.4	.8	1.2	Tbl 32 Q or F	B
C	Net Transfer Payments	[a]	.3	.8	.6	.7	.6	.8	GNP 115 M	C
D	Dispositions of Money	7.1	7.7	8.6	8.6	9.3	9.6	10.1	A+B+C+(G when G>zero)	D
E	Product Receipts minus Expenditures n e c [1]	6.9	7.7	8.4	8.6	9.2	9.6	10.1	GNP 114 T	E
F	Sources of Money	7.1	7.7	8.6	8.6	9.3	9.6	10.1	E+(−G when G<zero)	F
G	Discrepancy Included Above (Disposition)[2]	−.3	[a]	−.2	.4	−.1	.1	.4	P&B VI, Footnotes 1 and 2	G
	Banks and U S Monetary Funds									
H	GNP Expenditures	.2	.5	.3	.3	.4	.3	.2	GNP 117 G	H
J	Net Money Advanced or Returned to Others minus Net New Gold from Domestic Sources and Mint Purchases of Silver	.2	.1	0[a]	.1	.4	.3	.1	GNP 118 G	J
K	Net Transfer Payments	.1	.1	.1	.1	.1	.1	.1	GNP 118 E	K
L	Dispositions of Money	.5	.6	.5	.5	.9	.7	.5	H+J+K+(P when P>zero)	L
M	Product Receipts minus Expenditures n e c [1]	.5	.5	.5	.5	.5	.5	.5	GNP 117 T	M
N	Sources of Money	.5	.6	.5	.5	.9	.7	.5	M+(−P when P<zero)	N
P	Discrepancy Included Above (Disposition)[2]	0[a]	−.1	.1	0[a]	−.4	−.2	.2	P&B VII, Footnotes 1 and 2	P
	Private Insurance Carriers									
Q	GNP Expenditures	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	GNP 119 B	Q
R	Net Money Advanced or Returned to Others	1.9	1.9	1.8	1.9	2.3	2.6	3.0	Tbl 32 S+T	R
S	Dispositions of Money	1.9	2.0	2.0	1.9	2.3	2.7	3.0	Q+R+(W when W>zero)	S
T	Product Receipts minus Expenditures n e c [1]	.6	.7	.8	.8	.9	1.0	.9	GNP 119 P	T
U	Net Transfer Receipts	1.2	1.3	1.2	1.1	1.3	1.8	2.0	GNP 119 W	U
V	Sources of Money	1.9	2.0	2.0	1.9	2.3	2.7	3.0	T+U+(−W when W<zero)	V
W	Discrepancy Included Above (Disposition)[2]	−.1	.2	.2	0[a]	−.2	.1	−.1	P&B VIII and IX, Footnotes 1 and 2	W
	Security and Realty Firms *et al*									
X	GNP Expenditures	2.0	2.2	2.2	2.6	2.7	3.2	2.5	GNP 121 D	X
Y	Net Transfer Payments	.2	.2	.2	.2	.2	.3	.3	GNP 121 Z	Y
Z	Dispositions of Money	3.8	3.2	2.4	3.3	3.1	3.6	2.8	Y+Y+(d when d>zero)	Z
a	Product Receipts minus Expenditures n e c [1]	2.6	2.4	2.3	2.9	2.9	3.2	2.9	GNP 121 U	a
b	Net Money Obtained thru Financing	1.2	.8	−.1	.4	.2	.3	−.8	Tbl 32 U or H	b
c	Sources of Money	3.8	3.2	2.4	3.3	3.1	3.6	2.8	a+b+(−d when d<zero)	c
d	Discrepancy Included Above (Disposition)[2]	1.6	.7	−.1	.5	.2	.1	−.7	P&B X, Footnotes 1 and 2	d
	The Rest of the World									
e	GNP Expenditures	0[a]	.1	1.1	.9	1.7	1.3	.1	GNP 123 D	e
f	Dispositions of Money	0[a]	.1	1.1	.9	1.7	1.3	.1	e	f
g	Product Receipts minus Expenditures n e c [1]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	0[a]	GNP 123 F	g
h	Net Transfer Receipts	.2	.2	.2	.2	.2	.2	.3	GNP 123 I	h
i	Net Money Obtained thru Financing	−.2	0[a]	1.0	.7	1.5	1.1	−.2	Tbl 32 V or J	i
j	Sources of Money	0[a]	.1	1.1	.9	1.7	1.3	.1	g+h+i	j
	All Transactors									
k	Total GNP Expenditures	79.7	85.1	79.3	85.3	95.4	118.0	154.0	Tbl 33A (A+H+P+W+d)+ Tbl 33B (A+H+Q+X+e)	k
m	Product Receipts minus Expenditures n e c [1]	80.8	88.2	80.6	85.3	94.3	118.8	150.4	Tbl 33A (E+L+T+Z+g)+ Tbl 33B (E+M+T+a+g)	m
n	Discrepancy Included Above (Disposition)[2]	1.1	3.1	1.3	0[a]	−1.2	.8	−3.6	m minus k = Tbl 8 line Z minus GNP 117 E+F.	n

[1] Receipts from all product and service transactions (including business taxes) minus non GNP, nontransfer expenditures.

[2] A negative entry is regarded as a source of money not accounted for.

Note Due to rounding, figures for various lines calculated by formulas given in the source column may differ slightly from the entries shown.

Reproduced with permission from M. A. Copeland, *A Study of Moneyflows in the United States*, National Bureau of Economic Research, New York, 1952.

Line k of Table II B shows a growth of GNP expenditures for the entire period with the exception of the one year 1938. These figures do not agree with the GNP data published by the NID, but this is to be expected since the moneyflows data omit a host of accrual entries and transactions in kind. (The direction of change is nevertheless the same as in the NID data.)

As a first approximation we may agree that any transactor who reduces his loanfund balance in order to increase his GNP expenditures or reduces his GNP expenditures to increase his net financial claims is in some sense taking the initiative. An examination of Table II gives us an insight into what was happening in the period 1937–1942 In 1938 the major reductions in expenditures are found in the household account (line A) and the industrial corporations account (line P). An increase in expenditures by the federal government more or less offsets the other reductions. What the moneyflows account tells us that the GNP accounts do not is that households reduced their loanfund balances by almost one-half billion in order to sustain their expenditures, whereas industrial corporations and business proprietors retrenched sufficiently to add to their loanfund balances. We have here confirmation of initiative attributed to the business world

The role of corporations in the recovery of 1939 seems to have been fairly passive Decisive action can be attributed to state and local government (SLG) and business proprietors and partnerships (BPP). It should be noted that the increase in expenditures by the latter group is larger than that of the state and local governments, and that the reversal of behavior is more distinct. The BPP sectors increase in GNP expenditures occurred in spite of a sharp fall in the receipts recorded on line Z. There is a much larger element of passive responsiveness in the SLG activity We may note in passing that apparently it is not only the sign of the excess of receipts for expenditures that is analytically significant; the change in the magnitude of this excess requires comparable attention

By far the largest increase in GNP expenditures is found in the households (HH) account But in this account little evidence of initiative by this sector exists The federal government sustained its activity and raised transfer expenditures. Its major contribution to recovery thus seems to have been indirect. Among the other sectors there is little that is unusual.

The rest of the table confirms what has generally been known all along, namely, that the rapid growth of production during the early 40's was largely the consequence of the federal government's war pro-

gram. It was the only sector with a declining loanfund balance.

This examination is obviously cursory. Since its main purpose is to supply motivation for the study of moneyflows we have not examined many of the ramifications of Table II. Later in the paper we shall supplement this by utilizing the accounts to examine the financial ramifications of this ebb and flow of activity. In the meantime we return to a more careful examination of the moneyflows structure.

Sectoring

Copeland has chosen to use the so-called object or type of transaction basis for sectoring.[9] In this type of account the entries show the types of transactions for which funds were used and the types from which funds were acquired. We may note that the sectoring in *Moneyflows* substantially exhausts the economy The greater detailing of sectors in the moneyflows accounts is not just a breakdown of "parent" gross national product sectors. The classifications appropriate to the two sets of accounts are different. The setting up of gross national product accounts is plagued with the difficulties inherent in trying to force our institutions into two way moulds, those functional sectors that produce the gross national product (and pay distributive shares) and those that use up the gross national product (and receive distributive shares), i e., intermediate and ultimate sectors. In the moneyflows accounts the classification of a transactor is determined by his role in the credit markets as well as in the goods and service markets. With respect to the latter market the criterion seems to be more the type of product than its stage—ultimate or intermediate—in the production process. This must be so since sectoring by the latter criterion necessitates accounting vivisection of transactors whereas the consideration of financial flows necessitates keeping transactors intact. It should be clear, furthermore, that the introduction of such further criteria as financial activity will, in itself, lead to the multiplication of sectors if the degree of homogeneity of the sectors is not to be reduced. To a certain extent moneyflows sectors represent institutionally distinct entities important to understanding how a money economic system operates. In other societies, with other institutions, different sectoring may be required. It is much easier to export gross national product sectoring than moneyflows sectoring.[10]

Again, the gross national product accounts are oriented towards facilitating value judgments as well as descriptions of the economy as a

[9] *Ibid.*, p 101.

[10] However, see Dudley Seers, "The Role of National Income Estimates in Statistical Policy of an Underdeveloped Area," *The Review of Economic Studies*, 20 (1952–3), 159–162.

vast productive apparatus. The gross national product concept has developed from concept of national income The latter, in turn, was originally designed as a measure of national welfare.[11] As such it involves judgments of what is and what is not socially useful The accounts are used both for measuring welfare and for analyzing economic activity The moneyflows approach, on the other hand, is oriented only towards an analysis of what happens It views the economy as a pecuniary one with all the appropriate financial trappings as well as a productive apparatus. As a consequence, no sectors are identified as final in the moneyflows accounts

Since the gross national product accounts consolidate transactions between members of the intermediate sector, it makes no difference whether the industrial classification is by establishment or by ownership unit Since the moneyflows accounts are largely combined,[12] however, many moneyflows totals are not invariant with respect to the transactor unit and the ownership unit is taken as basic

On the whole, the moneyflows accounts are on a much grosser basis than are the gross national product accounts With some important exceptions, intrasector transactions are not consolidated In Copeland's study complete decision-making units are preserved intact insofar as possible and transactions between such units are recorded Among the exceptions are transactions in existing real estate, the accounts for the rest of the world, the federal government, and the banking sector Inadequate data necessitate using merely the net balance of real estate transactions[13] We have no choice but to consolidate when dealing with the rest of the world The federal government is considered a single transactor since it is considered a single decision-making unit. Thus inter-agency transactions are not recorded. And finally, the grounds for consolidation of the banking sector are "to bring out clearly the relation between banks and U S Monetary funds and the rest of the economy."[14]

To a certain extent these sectors are arbitrary The actual set of sectors with which we finally deal is a compromise between the analyti-

[11] See for example Pigou's use of this concept in the *Economics of Welfare*, 4th Edition, London, MacMillan 1932

[12] Copeland, *op. cit*, p. 125. A consolidated statement can be defined as "A statement showing [the] . . . condition or operating results of two or more associated enterprises as they would appear if they were one organization The preparation of a consolidated statement involves eliminations of inter-company accounts, investments, advances, sales, and other items " E L Kohler, *A Dictionary for Accountants*, Prentice-Hall, New York, 1952, p 98

[13] Copeland, *op cit*, p 87 Gross figures are now used in the extension of the moneyflows study now being conducted by the research staff of the Board of Governors of the Federal Reserve System See D. H. Brill, *et al., Progress Report on the Moneyflows Study* Washington 1951 (Mimeo)

[14] Copeland, *op. cit*, p 283

cally desirable and the statistically possible. We find ourselves dealing with the eleven sectors whose abbreviated titles head the columns of Table I To a certain extent Copeland's classification follows the lines of the *Standard Industrial Classification* [15]

Differences in sectoring principles make direct comparison of sectors rather difficult. When the sectoring is by the activity principle, the same transactor may be found in more than one account. In the gross national product Consolidated Business Income and Product Account we find some, but not all, transactions by members of several moneyflows sectors For example, we find that income generating, indirect tax, etc , expenditures, and capital consumption allowances of industrial corporations (Sector III) are included in the above gross national product account. On the other hand, capital expenditures by this same moneyflows sector are entered in the Gross Savings and Investment Account by the Department of Commerce However, in the case of business proprietors, etc., (Sector IV), the corresponding current expenditures by some members, such as the private non-profit organizations, are treated by the Department of Commerce as expenditures of the personal sector. The only moneyflows sectors in which all the members play a role in the Consolidated Business Income and Product Account are farms (Sector II), industrial corporations (Sector III), and the insurance carriers (Sectors VIII and IX). Thus government corporations whose income and product transactions are included in the NID account are wholly transferred to the appropriate moneyflows government sector.

Transaction Types

Copeland has set up fourteen types of transactions (the first fourteen row captions of Table I) The first thirteen types of transactions are collectively labelled ordinary transactions. Changes in loanfund balances[16] are shown on the table in detail. If the details are combined into a single net total, the result is what Copeland calls "money obtained through financing or net money advanced or returned to others " For each transactor this equals the difference between his ordinary dispositions and sources of money except for statistical discrepancies.

Generally speaking the titles of these transaction types are self explanatory. Among those that are not so self evident is customer money-

[15] Government Printing Office, Vol. 1, 1941, Vol 2, 1942 Copeland, *op cit*, p 37.

[16] Loanfund balances are the totality of financial claims owned and owed by a transactor or sector They include currency and deposits, book credit accounts, securities, mortgages, and in the case of monetary authorities monetary metals They do not include insurance policy reserves and other strictly accrual items.

flows (item 7). This type of transaction includes the purchases and sales of all non-financial goods and services not specifically included in the other transaction types It thus includes purchases and sales of second hand goods and intermediate goods as well as final goods. In the latter are included many capital goods.[17]

A second transaction type, the nature of which is not immediately obvious, is net owner takeouts. This is defined by Copeland as "cash withdrawals from proprietorship account by the owners of unincorporated businesses and farms and the lessors of real estate minus new money invested by these owners "[18] The households are the only recipients of this flow. As can be seen, the flow originates in three sectors. Logically, the computation of this account is equivalent to (1) combining the NID's income of unincorporated business and rental income of persons, (2) subtracting the components of these which are not considered moneyflows to households such as various accrual and imputed items, and (3) adding certain moneyflows to households not already included such as, for example, cash withdrawals of a proprietor in excess of the income earned by the business. The reconciliation of the moneyflows type of transaction with the two NID types is obviously anything but simple.

The final transaction type that calls for mention is public purpose payments Generally speaking these are transfer items Logically they should include gifts by households to households, but for fairly obvious practical reasons, the estimates do not Logically, this transaction type should include all international transfers, but actually it includes only personal remittances abroad.

To some extent, the inclusion or exclusion of particular transactions in the moneyflows accounts are predetermined by the condition that the accounts must balance and the prior decision to include or exclude other transactions. The introduction of net changes in the loanfund balances as a sort of balancing item in the account has important ramifications with respect to what can and can not be excluded. The counterflow of any flow which affects these net changes must necessarily be included or the balance will be destroyed. Thus, for instance, cash interest received must be included regardless of its economic significance. The exclusion of this item would make the difference between ordinary receipts and expenditures less than net changes in the loan-

[17] See below. Note also that construction is included in installments to contractors rather than in customer moneyflows. Note too that in the case of second-hand goods, only half the purchase cost is included.

[18] Copeland, *op. cit.*, p. 17.

fund balance which this difference is supposed to equal. On the other hand the sole consideration in the case of money-changer transactions is their significance, or rather in this case, their lack of significance. These transactions are entirely financial and affect both sides of each account equally. While for many purposes any transaction which merely affects the composition of the loanfund balances without affecting their size can be excluded, the information would be essential for most analysis of monetary and credit flows.

Moneyflows are neither on a strictly cash nor accrual basis.[19] Wages and taxes are on a cash rather than accrual basis.[20] But credit and installment purchases are recorded as well as cash purchases. However, since the system of accounts embraces only transactions between distinct parties and only transactions that involve the use of cash or other financial claims, neither imputed income, income in kind, nor internal transactions such as depreciation appear.[21]

The inclusion of various credit transactions involves the underlying and important institutional fact that in a good many transactions other financial claims are used instead of money. Trade credit is one such substitute A further consequence is that offset settlements are kept as distinct transactions.[22]

The various factors noted above contribute to the differentiation of the moneyflows and NID accounts. But perhaps one of the most striking differences between these two sets of accounts is the failure to identify capital formation items in the former. A glance at Table I will show this is so. This difference is not the result of inherent differences in the two sets of accounts. With the expansion of the number of sectors in the moneyflows accounts, it was statistically impossible, with the data available at the time, to identify sales that go on capital accounts.

After adjustment for the imputation and accrual items the relationship between the moneyflows accounts and the gross national product accounts can be displayed and tables such as Table II set up. However, the resulting accounts are deceivingly simple. While the conceptual relationship between the two systems is fairly straightforward, an actual statistical reconciliation of the two sets of accounts is exceedingly complex and difficult.

---

[19] P. 19.

[20] See S. J. Sigel, "A Comparison of the Structure of Three Social Accounting Systems," Conference on Research in Income and Wealth, October 1952, N B.E.R. (mimeo). Mr Sigel's remarks are directed at the revised system used at the Federal Reserve Board but most of his remarks apply as well to *Moneyflows*

[21] Copeland, *op. cit.*, p. 69.

[22] Copeland, *op. cit.*, p. 69.

The Structure

When a transactor participates in a main money circuit transaction, four simultaneous entries are recorded One pair of entries records the disposition and sources of money of one transactor. The other pair records the dispositions and source of the other party to each transaction. On Table I we can see what would have been the accounting effect if households had purchased an additional $100 million worth of goods from, for example, corner groceries, on a credit basis. In the household account (Sector 1) dispositions of money on account of customer moneyflows (item 7) would read $51 3 billion instead of $51 2 billion as it now does In the same sector account this change would be balanced by a change in book credit (L_+) (increased Accounts Payable) from $.2 billion to $ 3 billion Similarly in the account of Sector IV, to which corner groceries belong, sources on account of customer moneyflows would read $50 5 billion instead of $50 4 and book credit ($A_+$) (increased Accounts Receivable) would read $.4 billion instead of $ 3 billion It can be seen easily that this procedure leaves the accounts of both sectors in balance and at the same time leaves the rows of both customer moneyflows and book credit in balance

In the table, by examining the rows, we can see to which sectors the transactions have been sources and to which they have been dispositions of money, and of course the magnitude of the dispositions and sources of money on the account of the transactions Similarly, by examining the columns we can see the transactions of any particular sector. If the financial section, the portion of the table showing net increases or decreases in financial assets and liabilities, had been in terms of absolute values instead of changes, we would have had what Copeland calls statements of payments and balances A separate computation of loanfund financing, i e, net changes in balances of financial assets and debts, would be necessary to complete the statement

"Net money obtained through financing" and "net money advanced or returned to others" can be estimated either by balancing the ordinary receipts and expenditures or from the changes in the loanfund balances. The two estimates should be equal except for statistical discrepancies and deviations from accounting uniformity If we recapitulate the funds obtained and advanced, we can see which sectors advanced money and which sectors obtained it This is shown in Table III which Copeland labels "Net Money Obtained or Advanced a/c Loanfund Financing"

Table I shows the moneyflows for only one year. We can see the matrix of mutual constraints that exist in the economy by examining

the table in its totality. Neither the sources nor the dispositions of money of any sector can change without resulting in a corresponding

TABLE III

NET MONEY OBTAINED OR ADVANCED a/c LOAN-FUND FINANCING
(Millions of Dollars)

	1936	1937	1938	1939	1940	1941	1942
NET MONEY OBTAINED BY							
A Households	0	0	400	0	100	0	0
B Farms	0	300	100	0	200	0	0
C Industrial Corporations ..	0	1,100	0	0	0	0	0
D Business Proprietors and Partnerships et al	0	0[1]	0	300	0	0	0
E The Federal Government	5,050	300	1,350	2,200	2,400	10,050	41,150
F State and Oocal Governments	0	0	0	400	0	0	0
H Security and Realty Firms et al	1,200	800	0	400	200	300	0
J The Rest of the World	0	0[1]	1,000	700	1,500	1,100	0
K All Transactors. . .	6,200	2,500	2,800	4,100	4,500	11,500	41,200
NET MONEY ADVANCED OR RETURNED BY:							
L Households	2,300	100	0	300	0	5,000	19,600
M Farms	300	0	0	600	0	400	1,800
N Industrial Corporations	300	0	200	2,200	1,600	3,600	10,200
P Business Proprietors and Partnerships et al	700	0	800	0	500	300	4,300
Q State and Local Governments	300	200	300	0	400	800	1,200
R Banks and U S Monetary Funds	400	500	300	400	800	500	200
S Life Insurance Companies	1,500	1,400	1,600	1,600	1,700	2,200	3,400
T Other Insurance Carriers	400	400	200	300	600	400	600
U Security and Realty Firms et al .	0	0	100	0	0	0	800
V The Rest of the World	200	0	0	0	0	0	200
W All Transactors	6,500	2,600	3,500	5,400	5,600	13,300	41,200
X Discrepancy (Money Advanced or Returned minus Money Obtained)	300	100	800	1,300	1,100	1,800	0[2]

[1] Less than $50 million

[2] Lies between ± $50 million

Note Due to rounding columns may not precisely downtotal

Source The sources of the individual items appear in a final column in the original table, but since they are in code and refer to the extensive appendix of *Moneyflows*, they are not reproduced here.

Reproduced with permission from M. A. Copeland, *A Study of Moneyflows in the United States*, National Bureau of Economic Research, New York, 1952.

change in the account of some other sector. In other words no sector can change its expenditures, receipts, or composition of its loanfund balance without affecting one or more other sectors.

For many purposes it is better to arrange the data, as Copeland has, in separate tables for each sector, transaction, and loanfund balance and include the data for every year for which they are available.[22] It is then possible to get a historical picture of moneyflows, as in Tables II, and III. For most analytic purposes the time series variety of tables is to be preferred. For behavior analysis the structure form gives only a partial picture. The greatest advantage of Table I lies in demonstrating the mutual interdependence of our society.

By examining the financial sections of Copeland's statements of payments and balances (which we do not reproduce here for lack of space) we can complement the picture of economic activity portrayed above with an analysis of its financial ramification We shall not attempt to be anything but sketchy We noted that in 1938 H.H. reduced and industrial corporation increased their loanfund balances. What implications did this have to the volume and distribution of negotiable claims? We may note first of all that of a fairly sizeable increase in currency and deposits outstanding households received relatively little and industrial corporations a great deal However, the gain of $ 9 billion in currency and deposits by the latter turned out to be far greater than its net money obtained through financing and is reflected in the liquidation of accounts receivable and an increase in paid-in capital. We find a shuffling from government to other securities and an increase in interest bearing debt. While we find that most sectors shared in the increase in currency and deposits the increase in federal securities was largely picked up by the banking and life insurance sectors. Interestingly enough, this did not mean that the life insurance companies reduced their holdings of currency and deposits. Quite the contrary these holdings were increased The ordinary course of events yielded them a sufficient excess of receipts over expenditures to make a substantial purchase of other loans and securities as well

Of interest also is the significant change in the distribution of federal securities in the early 40's Without exception all sectors purchased securities Banks continued to hold about one-half of federal obligations The percentage held a) by households fell from about 1/4 to about 1/5 and b) by life insurance companies fell from about 1/9 in 1940 to about 1/13 in 1942 On the other hand industrial corporations holdings rose from about 1/45 to about 1/12.

---

[22] The tables in the *Progress Report* are roughly similar to those in *Moneyflows*, in that they are both of the time series variety However, the two sets of tables do not have uniform categories. Copeland does not present a detailed comprehensive picture of his accounts as is found in Table I. However, the structure of the Federal Reserve Board moneyflows accounts for 1947 has been prepared and used as part of a paper on the moneyflows study given by D H. Brill to the Econometric Society Meetings in December 1951, and the format of Table I is an adaption of this to Copeland's accounts.

Again, in Table III we can see the considerable increase in loanfund balances that had already taken place by the end of the first year of World War II. We could go further and examine the composition of these changes in fascinating detail but again the purposes of this excursion are to supply illustrations of the types of analysis moneyflows makes possible and, not least, motivation. Analysis, as such, of the period is beyond the scope of this paper.

Discrepancies

The discrepancies disclosed in Table I may appear to be too large for comfort. To a large extent this is due to the data available rather than to deficiencies in the system itself But some of the discrepancies are inherent in the accounts There are three major sources of non-uniformity in the accounts that contribute to these discrepancies. They are non-uniformity in timing, valuation, and classification.

First, for example, non-uniformity in timing results from the fact that the debtor's record of payment precedes in time the creditor's record of receipts. It takes time for payments to come through the mails. This discrepancy is also reflected in the national currency and deposit account so that the deposit holdings of the payee are understated and the banking sector reports more liabilities than there are apparent charges against it in the currency and deposit account.[24]

A second mail float, which Copeland fails to note,[25] tends to make receivables exceed payables. The float arises from the time lag involved in mailing invoices to the purchasers of goods and services. The discrepancy will also be reflected in some other account. If, for instance, the transaction involves a customer moneyflow, the dispositions in this account will tend to lag behind the sources

Lack of uniformity in valuation arises from different accounting practices which often lead debtors and creditors to value the same items differently. In the book credit account, receivables are listed as net and payables as gross.[26] In other words, the creditors' estimates of uncollectable accounts are already written off, but no debtor has reason to believe that this is the case with his account. It is such items as these that make up the valuation gains and losses at the bottom of the table. They also give rise to a tendency for estimates of payables to exceed receivables offsetting to some extent the opposite tendency resulting from non-uniformity in timing. They introduce difficulties in the com-

---

[24] Copeland, *op. cit.*, p 234.

[25] There is reason to believe that Copeland ignored this mail float because of the inadequacy of the data.

[26] Copeland, *op. cit.*, p. 153.

putation of financial flows The financial flows recorded in moneyflows accounts are composed only of those changes in the financial assets and liabilities that occur as a result of market transactions. Write-ups and write-downs must be removed from book increments and decrements to show those changes actually resulting from market transactions.

Finally, lack of uniformity in classification is found in the case of the purchase of gold from domestic producers by the banking sector. The domestic producers treat gold as a commodity and consequently credit customer moneyflows. On the other hand, the banks treat gold as a loanfund item and record no corresponding debit to customer moneyflows.[27] Similarly in the sale of securities by brokers to households, the households often treat the transaction as affecting only the composition of their loanfund balance, while the brokers take cognizance of the service changes involved

The examples of discrepancies we have noted are not exhaustive of all the discrepancies that can and do occur. The discrepancies that do exist and the inadequacies of the data combine to introduce an element of imprecision in the moneyflows accounts Fortunately, however, many of the discrepancies balance against each other

Prospects for Moneyflows

Whatever defects may be found in the accounts, Copeland's *Moneyflows* stands as a significant contribution to economics The value of the framework does not depend solely upon the validity or lack of validity of the discretionary hypothesis developed from it by Copeland More than one model is compatible with the moneyflows framework in the same way as more than one model is compatible with the gross national product framework. One of the great advantages of the NID accounts is the ordering they make possible of the many series of data which we have had for a long time, and not incidentally, for the gaps in our statistical knowledge that they disclose The moneyflows accounts do the same for other series of data. They enable us to bring together into an integrated picture the financial and nonfinancial activities of the various major sectors

*The financial facts in the moneyflows accounts are financial market facts They record sales less purchases of stocks and bonds, of government securities, and other negotiable instruments not separately recorded in the accounts. As we have seen they can tell us who liquidated

---

[27] *Ibid.*, p 87.

* The following three paragraphs were written by Prof. J C Dawson of the University of Maryland. They appear substantially in their original form However, I have made a few changes and the responsibility for any error is, therefore, mine.

portfolios, who increased their debts, who made more use of trade credit, who borrowed from banks, and who drew down their cash balances We cannot find these financial market phenomena in the NID Gross Savings and Investment Account The latter does not tell us what financial forms personal saving takes, or what financial means were used by business or government to obtain funds It does not disclose the linkage between the two through our financial institutions.

Lest the reader find this unnecessarily vague let him attempt answers to the following by examining the Commerce accounts: How much money does non-corporate business obtain from the financial markets and in what forms? What impact would choking off bank credit have on outside corporate financing? What sectors do insurance policy reserves finance? How do we relate purchases of new houses and other durable consumer goods to the means of financing them, e g , the availability of consumer credit? It will be recalled that increased corporate security holdings by households, increments in insurance policy reserves, and purchases of new homes by prospective owner occupants are all components of personal saving—though not separately itemized in the Commerce residual estimate It may be said that the gross savings and investment account was not designed to answer such questions Yet much current economic analysis relating finance to production—analysis that ought to concern itself with questions such as these—is done in terms of this account

While some of the financial market transactions listed above are not separately identified even in the moneyflows account, most of them can be revealed by the various sector statements of payments and balances And via this account each sector's financial operations can be related to its operations in the production area, enabling a synthesis of these two major areas in over-all economic adjustment But part of the advantage is not that moneyflows accounts are sufficiently detailed for analysis, but that they are capable of being expanded further to show further details

Thus the usefulness of the accounts can hardly be questioned. We cannot and will not attempt to catalogue further the myriad of applications that are possible We may note that among other things, they provide us with an X-Ray into the fabric of the veil of money The confusion of the flows of goods, services, and resources with flows of money has led many an economist into errors that can be avoided by the utilization of the moneyflows framework or some variant thereof. The questions propounded by Copeland, some of which we have noted earlier, and his answers to them are impressive.

Furthermore, the accounts provide us with a framework with which we may be able to elaborate, or even reconstruct, existing theory, and mould theory into a more powerful instrument for analyzing current activity and providing greater insight into subsequent events. In the account of each sector there are possibilities of finding expenditure functions. Until the interrelationships between the various items of the accounts have been carefully examined, there exists the possibility that the mine of information that may exist in these accounts has not been exhausted.

A final word about the use of the word structure in this paper is in order. As has been noted, a structure of moneyflows is not exactly analogous to a structure of input-output relationships The data that appear in an input-output table are the results of economic organisms operating within the confines of technological or engineering relationships It is believed that the ratios of the values in the cells of the latter table are free to move only within narrowly confined limits As a consequence, the table for one bench mark year supplies us with a considerable amount of information about these relationships The same is not true in the case of moneyflows. Nor, incidentally, is it true of the gross national product accounts In the case of moneyflows, the ratios of the values in the cells are determined, to some extent at least, by behavior relationships. These, as we have discovered to our chagrin, are notoriously hard to establish It is a rash economist indeed that propounds a behavior relationship on the basis of data for one year. Even with data for seven years, Copeland was justifiably hesitant about propounding the existence of specific relationships Historical series, and long historical series, are a necessity for the investigation of such relationships Furthermore, whereas engineers can supply us with information about changes in technological relationships, we are unlikely to discover analogous changes in behavior relationships as easily. Consequently, there is a sense in which it seems that input-output studies are more efficient, though not necessarily more useful. Any such difference in efficiency, however, is not inherent in the nature of the two approaches, but rather in the subject matter with which they deal, and in the light of the difficulties that are now being encountered in carrying the input-output studies further, it seems that the difference in efficiency can easily be overrated.

THE MEASUREMENT OF SEASONAL MOVEMENTS IN PRICE AND QUANTITY INDEXES

BRUCE D MUDGETT
Brown University

THE problem of constructing a price index (or quantity index) that will correctly measure price or quantity changes over the seasons has occupied index number makers for many years, but no solution of this problem that has yet been proposed has been very satisfactory.

One method used has been to weight the index with weights appropriate to the *yearly* importance of the individual commodities and then to neglect variations in importance over the year. For commodities that disappear wholly from the market at certain seasons of the year the index has been carried forward over this period by retaining the last quoted price until a new quotation appears in the market. This is a typical procedure though there have been variations in detail It is not difficult to see that this is no solution of the problem at all, that, instead of measuring the actual price variations over the seasons the index number maker has introduced quotations that have nothing to do with particular seasons at all Consider an index, for example, that is published at monthly intervals and suppose that the weights for all commodities in the index are annual weights, that is, they are measures of the yearly, not monthly, importance of each commodity. Consider now the February index, on the not unrealistic assumption that there has been no price quotation for a given commodity since the previous November. The position taken here is that a monthly index for February constructed by including this commodity with a price belonging to the previous November and a weight measuring annual and not monthly importance, gives a resulting index with an element, so far as this particular commodity is concerned, that is wholly unrealistic. If indexes are measures, more or less exact to be sure, of actual price situations through which an economy has passed, then the procedure described deserves a rather solid condemnation as soon as an improve procedure can be devised.

There is described herewith a procedure for measuring seasonal price influences[1] that the writer believes to possess enough merit to justify publication and an invitation for critical appraisal by other statisticians.[2]

[1] The procedure will be developed in terms of price indexes, but it is directly applicable to quantity indexes with nothing more than the interchange of the p's and the q's of ordinary index number notation.

[2] A word needs to be said as to the origin of this solution. Over the period of the revision of the in-

Consider the concept of a yearly price level as measured by the traditional value aggregate $\sum p_0 q_a$ familiar in modern index number notation, where the summation is over the commodity list upon which the index is based and where the weights, q, are appropriate measures of the importance of the prices, p. The base period prices here are the p's and the weights the q's No question need be raised about the appropriateness of the q's, since they are completely general at this point and therefore any index number maker may give them specific content with respect to the way he defines appropriateness of weights. The above aggregative for the year 0 is built up by the continuous summation of value items over the year It can therefore be broken down into subaggregates, by months for example, if months are the smallest time intervals for which indexes are wanted If the year's aggregative represented a year's family expenditures based upon the annual consumption quantities, q_a, then its monthly components would measure the spread of the yearly total over the seasons.

A realistic measurement of price changes over the seasons can be built upon this conception The proposed index will first be designed using a fixed weight aggregative formula, but it will be shown later that the principle on which the measurement is based is independent of this formula and is equally applicable to any of the theoretically best formulas This generality is a matter of prime significance. In order accurately to describe this index an exact notation is required through which various needed aggregates are given meaningful and precise definition. This notation and the appropriate aggregates follow at once.

Notation

(1) p, q = prices, quantities of individual commodities.
(2) P, Q = indexes of prices, quantities.
(3) Superscripts, t, represent commodities They will be omitted where, in the context, they are not needed.
(4) Subscripts, ij, represent·
 years $i = 0, 1, 2, \cdots$
 months $j = 1, 2, \quad 12$.

dexes of wholesale and consumer prices of the United States Bureau of Labor Statistics, November 1949 to February 1953, many of the discussions which took place between the Bureau and the Technical Advisory Committee of the American Statistical Association were devoted to the discussion of the problem of seasonal measurement During these years several memoranda on the subject were prepared, significantly those by Doris Rothwell of the Bureau staff and by Abner Hurwitz, Chief of the Cost of Living Section of the Bureau staff The development of the present writer's views were influenced in an important way by these memoranda and by the discussions of them The writer submitted one memorandum which is the substance of what is presented here and wishes to say that while he is the sole author of what follows and takes full responsibility for it, it is nevertheless true that his own views on seasonal measurement were developed and his understanding of the subject greatly clarified by these discussions.

In general
0 = base year
i = a given year
ij = a given month of a given year.
(5) Number of commodities, t.
N_{aj} = number of commodities in list for month j (of the year a)
N_a = number of commodities in the list for the year a (some weight period),

$$= \sum_{j=1}^{12} N_{aj} \text{ minus duplications.}$$

Fixed monthly weights for commodity t·
$q_j^{(t)}$ = quantity of commodity t for month j.
$t = 1, 2, \cdots N.$
$j = 1, 2, \cdots 12.$
(In general, this magnitude is not the actual quantity for a given month obtained in the market for that month, but the fixed quantity used as weight.)

$$\sum_{j=1}^{12} q_j^{(t)} = q_a^{(t)} = \text{annual quantity (weight) of commodity } t.$$

Monthly average price, $p_{ij}^{(t)}$, and yearly average price, $p_i^{(t)}$, of commodity t for year i

Then, for $i=0$,

$$\frac{\sum_{j=1}^{12} p_{0j}^{(t)} q_{0j}^{(t)}}{\sum_{j=1}^{12} q_{0j}^{(t)}} = p_0^{(t)} = \text{base year average price of commodity } t.$$

And since

$$\sum_{j=1}^{12} q_{0j}^{(t)} = q_0^{(t)},$$

therefore

$$\sum_{j=1}^{12} p_{0j}^{(t)} q_{0j}^{(t)} = p_0^{(t)} \bar{q}_0^{(t)}.$$

Similarly for any i

Monthly and yearly aggregates and index numbers

Base year, 0

Month	*Monthly Aggregates*	*Cumulative Monthly Aggregates*
Jan.	$\sum_{t=1}^{N_{a1}} p_0 q_1$	$\sum_{t=1}^{N_{a1}} p_0 q_1 = A_{01}$
Febr.	$\sum_{t=1}^{N_{a2}} p_0 q_2$	$\sum_{j=1}^{2} \sum_{t=1}^{N_{aj}} p_0 q_j = A_{02}$
.	.	.
.	.	.
.	.	.
Dec	$\sum_{t=1}^{N_{a,12}} p_0 q_{12}$	$\sum_{j=1}^{12} \sum_{t=1}^{N_{aj}} p_0 q_j = A_{0\,12} = A_0 = \sum_{t=1}^{N_a} p_0 q_a$
Year	$\sum_{j=1}^{12} \sum_{t=1}^{N_{aj}} p_0 q_j =$	$\sum_{t=1}^{N_a} p_0 q_a = A_0$

Given year, i $[i=0, 1, 2, \cdots]$

Month	*Monthly Aggregates*	*Cumulative Monthly Aggregates*	*Index Numbers*
Jan.	$\sum_{t=1}^{N_{a1}} p_{i1} q_1$	$\sum_{t=1}^{N_{a1}} p_{i1} q_1 = A_{i1}$	$P_{0\,i1} = \dfrac{A_{i1}}{A_{01}}$
Febr.	$\sum_{t=1}^{N_{a2}} p_{i2} q_2$	$\sum_{j=1}^{2} \sum_{t=1}^{N_{aj}} p_{ij} q_j = A_{i2}$	$P_{0\,i2} = \dfrac{A_{i2}}{A_{02}}$
.	.	.	.
Dec.	$\sum_{t=1}^{N_{a\,12}} p_{i\,12} q_{12}$	$\sum_{j=1}^{12} \sum_{t=1}^{N_{aj}} p_{ij} q_j = A_{i\,12} = A_i$	$P_{0\,i(12)} = \dfrac{A_i}{A_0}$
Year	$\sum_{j=1}^{12} \sum_{t=1}^{N_{aj}} p_{ij} q_j =$	$\sum_{t=1}^{N_a} p_i q_a = A_i$	$P_{0i} = \dfrac{A_i}{A_0}$

From the index numbers given above for the year i it will be seen that a full set of all monthly indexes is available, as well as all yearly indexes Each monthly index however measures the cumulative price change of the given year through the month in question. That is, the February index measures the cumulative effect of price changes through January and February. Similarly for the other months. Several important properties of this index may be pointed out:

(*1*) The basic price index is the yearly index

(2) The monthly indexes within any year measure the change in the cumulating influence of the seasons. The December index for any year is identical with the year index.

(3) The price index for any month of any year upon any other month of that or another year may be obtained by direct division of the two indexes Thus the July i index on the June i base becomes

$$P_{0\,i7} \div P_{0\,i6}.$$

(4) The above procedure is independent of the formula used, in the sense indicated herewith:

(a) As used above, the price change is measured by the fixed weight aggregative formula.

(b) The procedure will give a measurement of price changes by the Laspeyres or Paasche formula, merely by replacing weights $q_j^{(t)}$ by $q_{0j}^{(t)}$ (Laspeyres) or by $q_{ij}^{(t)}$ (Paasche) Then any cross between these two is easily obtained The Marshall-Edgeworth formula is obtained by using $(q_{0j}^{(t)}+q_{ij}^{(t)})$ as weights.

(c) The question of using a fixed-base formula or at chaining consecutive yearly links is entirely independent of the procedure of measuring the monthly changes.

(d) One important operational result is inherent in this procedure, namely that the year index, upon whatever base, emerges as an end result from the calculation of the twelve monthly indexes

It is important not to lose sight of the significance of the statements (1) and (2) above The basic index is a yearly index and as a price or quantity index is of the same sort as those about which books and pamphlets have been written in quantity over the years. The within-the-year indexes partake of a slightly different character, in that they measure the cumulating influences of price and quantity changes over the seasons. But they do this with complete realism in the sense of giving a measurement of the price change or quantity change that actually took place This statement is not intended to refer to the use of approximate for exact weights (the problem of correct weights being a separate problem in index number theory and practice) nor to errors in price quotations (the collection and compilation of prices being a second separate problem in index number theory and practice). It is a proposal for replacement of a practice, which can never be anything but incorrect, of using annual weights, say, for February in place of Feb-

ruary weights, and of using, say, November prices for February when no price ever existed for February To introduce these non-appropriate and non-existent data into an index number in order to satisfy a public demand for monthly indexes is to create something out of nothing; and that it introduces a spurious element into index-number construction can scarcely be denied

The annual index here proposed is what an index has always been, namely a measure of price change upon a basis selected by the maker and measuring price change with an accuracy which is dependent on (1) the accuracy of the basic data, (2) the accuracy of the sampling involved, and (3) the efficiency of the formula used The monthly indexes, subject to all the elements of accuracy or of error enumerated above, can give an accurate measure of the cumulative influences of price change (or quantity change) throughout the months of the year, compared to the corresponding months of the year chosen as base; and this is done with the complete realism that is associated with the disappearance of some commodities at some seasons and their reappearance at others This sort of measurement has never yet been made, or at least the present author has never seen a description of it in the literature, but the price measurement that is here under consideration, it is submitted, is just what is wanted for a realistic settlement of issues that arise today between the proponents of various sides of modern controversies over price changes and over changes in real production

AN APPLICATION OF MARKOV PROCESSES TO THE STUDY OF THE EPIDEMIOLOGY OF MENTAL DISEASE

ANDREW W. MARSHALL AND HERBERT GOLDHAMER
The Rand Corporation

I. INTRODUCTION

WE SHALL present in this paper some methods that we have been applying to studies of mental disease These studies are epidemiological in character, that is they deal with the distribution and frequency of mental disease in different population groups We are primarily interested in getting answers to several substantive problems. In this paper, however, our concern is methodological and the emphasis is on the applied mathematical statistics and model construction of simple Markov processes The models that we shall discuss are very simple ones and in their present form are intended only to provide rough estimates of certain epidemiological parameters that are not directly observable or that can be secured only by expensive and time-consuming field surveys.

II. OBJECTIVES

The work that we shall report on has grown out of an interest in characterizing the age distributions of the mentally-ill population. These constitute the basic data of most epidemiological analyses. Age distributions are likely, however, to have a different significance according to the point in the pathological process used to represent the age structure of the population Thus patients may be distributed by the age at which the disease had its onset, by the age at which they sought medical care or were institutionalized or by age at death, to mention only a few possibilities The age distributions available in the literature of mental disease are mostly based on the age at which patients are admitted to a mental hospital. This is due, of course, to the ready availability of such data in hospital records and to the difficulty of specifying age of onset in the case of mental diseases whose early development is often insidious and in any case not usually investigated or recorded with care The widespread use of admission age has led in some cases to interpretations that are misleading. Thus one writer using hospital admission data argues that involutional psychoses in women are more closely associated with social than biological factors since their greatest incidence occurs after the involutional period. This

writer has developed a theory about the special stresses of life for women in the age period of about 50–55. However when age-at-onset data are used, the peak incidence for the involutional pychoses among women is approximately 45 years

There are a number of hypotheses and clues in the study of mental disease that are closely linked to the age characteristics of the disease process. For this reason we have thought it worth while to attempt to substitute for the prevailing age-at-admission distributions a series of distributions based on age-at-onset. We found that the mental hospital systems of Ohio and Illinois record (but have not tabulated) for each admitted patient an age-at-onset of the disease.[1] We recognize the deficiencies of these data and the numerous sources of inaccuracy by which they are affected We believe that the data are subject to a systematic bias that tends to displace the true date of onset forward toward the time of admission [2] If this is so, these age-at-onset distributions permit us to re-analyze the age structure of the patient population in terms of distributions that have at least been shifted in the right direction and represent the least amount of correction that is required.

The availability of age-at-onset data in Illinois and Ohio provided the possibility of attacking at the same time a closely related problem which concerned us It is this problem and the method of dealing with it that we now wish to discuss.

Important theoretical and practical considerations hinge on adequate answers to the question. What is the age-specific incidence of mental disease? This question has been answered primarily in terms of hospital first admission data, but rates obtained in this manner are significant for the incidence problem only if we can assume that the number of psychotics never hospitalized is negligible and that the lag between onset and hospitalization is very short Neither of these assumptions is obviously justified a priori; in any case the legitimacy of using hospital admission rates could only be established by comparing them with the true incidence rates

The most obvious means of securing a true incidence rate is by direct field investigation of a population. Field investigations must usually confine themselves to rather limited population groups and for this reason it is generally easier in these cases to secure a prevalence than an incidence measure. (By a prevalence measure we mean the proportion

[1] We wish to acknowledge our indebtedness and express our gratitude to the Ohio and Illinois Departments of Public Welfare for permitting us to reproduce their punch cards on which onset data are recorded.

[2] This presumption is supported by data provided by Larsson and Sjögren in a study of a Swedish population which has just come to hand. Cf. reference [5] and footnote 4

of a population that is affected at a given time; by an incidence measure, the proportion that has become affected for the first time in a given period, usually one year.) Prevalence measures can, however, be converted to incidence measures, although this will usually lead to some loss of accuracy in the latter. But whether one aims at prevalence or incidence measures, field survey procedures are costly and time-consuming. More important, it is extremely difficult to ensure that all non-institutionalized cases will be uncovered. Nonetheless, field surveys have been made with varying success. European, especially Scandinavian, investigators have contributed most to this work. The latest in a series of excellent studies are by Fremming [2], and by Larsson and Sjörgen [5]. In the United States we had for some time only two field surveys of value [4, 9], although others are now underway [3].

Our own approach to this problem has been quite different, although its adequacy will undoubtedly require testing in the light of rates established by direct field investigation. We have devised several very simple models of the process (or perhaps more accurately, the stages or states) involved in the passage from sanity to insanity, hospitalization and death Using the assumptions involved in one of the models and the data that the model requires, we have attempted to estimate what the total incidence rate must have been in order for it to have generated the known hospital admission rates. As we shall see more fully later, if we know something about the age-specific death rates of non-institutionalized psychotics, the death rates of the normal population, hospital admission rates and the lapse of time occurring between onset of the disease and hospitalization for individual patients, we can very likely infer what the total incidence in the population must have been.

Deficiencies in the age-of-onset data and the absence of reliable data on the death rates of non-institutionalized psychotics would make it difficult at this stage to arrive at more than rough order of magnitude estimates of the total incidence rate even assuming the adequacy of the models that we are using Nonetheless, we feel that the development of such models is important so that we can take advantage of more reliable data when they become available and in order to stimulate a greater interest among epidemiologists in securing such data.

Our interest in model construction is not confined to the immediate aim of securing estimates of the total incidence rate These models, simple as they may be, provide a partial picture of the underlying process that generates a given incidence rate. They thus enable us to see more clearly the role of several factors in producing these rates and to

establish the relative sensitivity of the rates to these factors. In this respect the models are more illuminating than an incidence measure secured by direct field investigation. To be sure, the models that we discuss here may not be of very great interest since their parameters do not represent factors that are usually thought of as causes of insanity. The models do, however, enable us to see how the total incidence rate is related to such parameters as death rates and the lag between onset and admission The models also permit us to deal with certain questions in a more unified manner Thus incidence and prevalence measures are sometimes treated as rather distinct measures whose relationships are not clearly defined In the model framework the incidence measure is clearly seen as specifying the frequency with which certain transitions occur between selected states of the model, and prevalency as specifying the number of persons who at any one time are in certain selected states of the model.

We now turn to an examination of the models themselves.

III THREE MODELS OF THE ONSET-ADMISSION PROCESS

The three models we will present are very simple The stochastic process used is Markov and only a few states are assumed. The models treat only the process leading up to first admission to a mental hospital and therefore do not consider discharge and readmission This is, of course, a usual focus for studies of mental disease. The models are similar to those suggested by Neyman in connection with the post discharge histories of tuberculosis patients [6]

Model I:

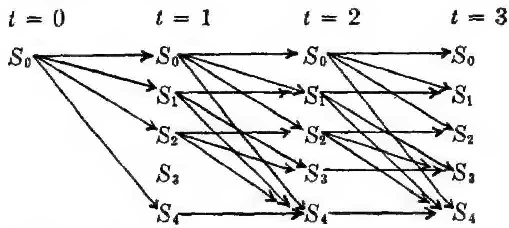
Our first model of the onset-admission process, Model I, makes use of the following states·

$S_0(t)$ = alive, sane
$S_1(t)$ = alive, insane (mild), unhospitalized
$S_2(t)$ = alive, insane (severe), unhospitalized
$S_3(t)$ = insane, hospitalized
$S_4(t)$ = dead, outside of mental institution (hospital)

State $S_4(t)$ refers only to deaths outside of institutions since after admission to a hospital subsequent changes of status are not of interest to the model Similarly, we can assume that persons entering hospitals never leave without altering the adequacy of our model for the analysis of the onset-admission process for first admission cases. Thus death and hospitalization are considered to be absorbing states, i.e , states from which there is no return It is easily seen that this is perfectly proper

from the point of view of the process we wish to study since no one can be a first admission for a second time, just as he cannot die twice. In this way sub-processes can be factored out for study from larger processes and made complete by suitable conventions concerning certain of the states and transition probabilities.

If we consider time intervals short enough so that only one transition to a different state can occur within it, the process of Model I may be diagrammed as follows:



The arrows indicate the possible changes of state that may be accomplished in one step In addition, we may characterize our model by a matrix of transition probabilities p_{ij} where p_{ij} is the probability of going from S_i to S_j in the small basic time interval we have chosen. Thus, for Model I the matrix P of transition probabilities is

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} & 0 & p_{04} \\ 0 & p_{11} & 0 & p_{13} & p_{14} \\ 0 & 0 & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

where $\Sigma_j p_{ij} = 1$ for all i; i.e , the sum of the probabilities in any row must equal one

We may differentiate two levels of assumptions required to justify the application of models of this general type to the currently available onset data These data, for each individual admitted to the hospital, consist of: (*1*) The age of admission, (*2*) The length of time between first onset of the disease and admission in units of one year, and (*3*) the calendar year of admission [3] Assumptions of the first kind concern the number of states in the model to approximate the real world and the possibility or impossibility of reaching one state from another directly

[3] The data we have contain more information (finer time units) about time lags of less than one year between onset and admission and also give month, day, and year of admission. For various reasons, however, we are unable to make use of this additional information.

in one step. In Model I for example this kind of assumption has to do with the inclusion of two states of insanity and assumptions that no one ever regains sanity once insane, no one becomes a severe case if he at any time is a mild case or vice versa. It is also assumed that the population does not change in any other way, for example, by immigration. These types of assumptions determine the size of the matrix and the number and position of the zeros in the matrix. The second kind of assumption has to do with the structure of the model itself, principally with the assumption of constant transition probabilities over time. In practice this must refer to both calendar and age time for we will have to group together the data for several years and several age groups. As is usual in empirical work we will have to be content to claim the model as a first approximation. As will be indicated later these claims can be checked to some extent by statistical means.

In the analysis of Model I one of the first problems is to evaluate P^n, the nth power of the matrix of transition probabilities In addition to the evaluation of P^n, which would be sufficient to solve the usual problem involved in the analysis of models of this sort, we must notice that our problem is rather special. This can be seen as follows: Let us suppose we have a cohort of persons of age x. We are interested in the rate (ultimately the probability of going insane at age x) at which the individuals in this group make the transition from S_0 to S_1 or S_2 during the time period of one year, or during age x to $x+1$. Our data consist entirely of persons who enter S_3, some perhaps as late as $x+2$, $x+3$, etc. Thus for instance, we have to evaluate the probability of being in S_1 or S_2 at the beginning of age $x+1$ and arriving in the hospital at age $x+1$, $x+2$, $x+3$, etc. For this purpose we break the process into two parts and define

$$P = \begin{pmatrix} p_{00} & p_{01} & p_{02} & 0 & p_{04} \\ 0 & p_{11} & 0 & p_{13} & p_{14} \\ 0 & 0 & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

to characterize the process during age x and

$$P_{00} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & p_{11} & 0 & p_{13} & p_{14} \\ 0 & 0 & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

to describe the process from age $x+1$ on. Here we will not go through the algebra involved in evaluating p^N, $P^N P_{00}{}^N$, $P^N P_{00}{}^{2N}$, $\cdots$, etc.

Our procedure has been to evaluate the above expressions for the time-discrete process and then to replace this by the continuous process which is its limit as the basic time intervals go to zero. The final expressions are then written in terms of a time parameter t and λ_{01}, λ_{02}, λ_{04}, λ_{13}, λ_{14}, λ_{23}, instantaneous rates of change between states S_i and S_j.

Model II:

As has been seen above, Model I may be designated as a two insanity state model; similarly Model II may be designated as a one insanity state model with recovery and relapse. In this model the possible states are:

$S_0(t)$ = alive, sane
$S_1(t)$ = alive, recovered from insanity, never admitted to hospital
$S_2(t)$ = alive, insane, unhospitalized
$S_3(t)$ = insane, hospitalized
$S_4(t)$ = dead outside hospital

As before, S_3 and S_4 are absorbing states since we are concerned only with the process leading to first admissions. The inclusion of a separate state S_1, recovered, gives us a flexibility that would not be available if we had treated recovery as a transition from S_2 to S_0. This simplification can be introduced later but for the present we will allow for the possibility that p_{02}, the probability of first onset, may not be equal to p_{12}, the probability of relapse Again as in Model I we define the matrices of transition probabilities.

$$P_L = \begin{pmatrix} p_{00} & 0 & p_{02} & 0 & p_{04} \\ 0 & p_{11} & p_{12} & 0 & p_{14} \\ 0 & p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Model III:

Model III is a modification of Model I, having two insanity states with a non-reversible transition from the mild to the severe state. The possible states are as in Model I and the matrices of transition probabilities are

$$P = \begin{bmatrix} p_{00} & p_{01} & 0 & 0 & p_{04} \\ 0 & p_{11} & p_{12} & p_{13} & p_{14} \\ 0 & 0 & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Thus no direct transition from sanity, S_0, to a severe state of insanity, S_2, is possible. A mild case of insanity may become severe, but not the reverse Death, S_4, and hospitalization, S_3, are, as before, absorbing states.

As can be seen, all of the above models have the same dimensionality; that is, they all have the same number of possible transitions In these terms they are models of equivalent complexity although just which transitions are allowed is very important in determining the complexity of the possible individual case histories Some exploratory work indicates that models of at least this dimensionality are required to fit the data we have A model of the same basic type as Model I but having only one state of insanity has been tried This model which we may designate as Model I′ is the same as Model I with states S_1 and S_2 combined When applied to our data, it does not appear capable of producing a good fit The data seem to require a model which will satisfy two conditions: (*1*) a large proportion of the persons of a given age having an onset of mental disease are admitted to a mental institution within one year of the onset of the disease and (*2*) the frequency distribution of the elapsed time between onset and admission has a long tail These conditions obviously suggest trying something like Model I with its two insane groups, one S_1, with long time lags between onset and admission, and the other, S_2, with short time lags between onset and admission Model II also has the possibility of reproducing the long tail on the elapsed time distribution. Without some changes Model III does not appear to be as adequate in this regard

IV PROBLEMS IN THE APPLICATION OF THE MODELS

We turn now to a discussion of the problems of using these models for the accomplishment of our objectives: (*1*) Deriving age-specific rates of onset of mental disease from admission data, and (*2*) Constructing prevalency estimates from admission data. First, it is obvious that at best, even if we ignore the clear inadequacies of currently available data, only approximate results are possible A really adequate stochastic process would, in addition to having a generally higher level of de-

tail, have transition probabilities which were, even locally, functions of age. With our constant transition probability models the best we can hope to do is locally to approximate the true process in five year age groups. We will return to this point. Second, a characteristic of our data, which is decisive with respect to certain details in the estimation of unknown parameters in the models, is what may be called its cross-sectional character. Rather than having the data pertinent to one cohort followed over several time periods, we have data on several cohorts at one time period. This point can be made more clear as follows: Let $N(t)$ be a vector describing the state of a cohort of N individuals at time t. In the case of our models the vector would have 5 elements. Suppose we observed the cohort at times $t_0, t_1, t_2, \cdots$, then our estimation procedures would be based upon the conditional probabilities $P[N(t_i+1)|N(t_i)]$. This is the usual form that estimation problems take in Markov processes In contrast we have observations on a series of cohorts $N_1(t), N_2(t), \cdots$ at time t_0, say.

As an example of how the estimation of the unknown parameters in one of our models might be treated, we will sketch out the estimation procedure for Model I in a typical case Let us assume that adequate values are available for λ_{04}, λ_{14}, and λ_{24} This is true for λ_{04}, which is essentially the death rate q_x in a suitable life table, and studies of the death rates of hospitalized patients shed some light on the possible values of λ_{14} and λ_{24} In any case the latter can always be treated as parameters of variations in the study and computations performed for bracketing values Thus we are left with four parameters, λ_{01}, λ_{02}, λ_{13}, and λ_{23}, to estimate We will require then at least four equations relating these parameters to the data for this purpose In the particular case at hand we used exactly four equations in order to simplify the computation A single iterative method of solution was used starting with initial values obtained from nearby age groups We also arranged it so that we have some observations left over in order that we may have an independent estimate of the goodness of fit of our estimated model to the whole of the data This is essentially an estimation procedure based upon the method of moments and yields consistent estimates of the parameters An alternative procedure would be to use Neyman's minimum χ^2 methods of estimation [7], using five or six equations, but for equations of the sort implied by our models this method would require considerably more computational effort.

Denote by $P_{03}(t)$ the probability of an individual being in S_3 by time t if he began in S_0 at time zero, where time zero is defined relative to age X. Writing out the probabilities of an individual arriving in S_3 for

the first time during the yearly time periods (0, 1), (1, 3), (3, 5), and (5, 7) as functions of λ_{01}, λ_{02}, λ_{13}, and λ_{23}, we have [for the exact expressions see appendix]

$$
\begin{aligned}
f_1(\lambda_{01}, \lambda_{02}, \lambda_{13}, \lambda_{23}) &= \{P_{03}(t = 1)\} = P_1\\
f_2(\quad \cdots \quad) &= \{P_{03}(t = 3) - P_{03}(t = 1)\} = P_2\\
f_3(\quad \cdots \quad) &= \{P_{03}(t = 5) - P_{03}(t = 3)\} = P_3\\
f_4(\quad \cdots \quad) &= \{P_{03}(t = 7) - P_{03}(t = 5)\} = P_4
\end{aligned}
$$

P_1 is the probability of being admitted to a hospital within one year of onset, P_2 is the probability of being admitted within three years of onset but more than one year, etc. All of these computations are specific to one age or age group; in our work we have used five year age groups. All of the parameters are different for each age group; the death rates being our best guess based upon the information at our disposal, the other parameters are then estimated from the data. This means that we end up with a model with fixed parameters being fitted as a piecewise approximation to the true underlying process where the parameters must be considered as continuous functions of age.

Thus from the Ohio data, for males having onsets between the ages of 35 and 39, for example, we have estimates of the probabilities $P_1, \cdots, P_4$, denoted by $\widehat{P}_1$, $\widehat{P}_2$, $\widehat{P}_3$, and $\widehat{P}_4$ The problem then is to solve the system of equations

$$
\begin{aligned}
f_1(\lambda_{01}, \lambda_{02}, \lambda_{13}, \lambda_{23}) &= \widehat{P}_1\\
f_2(\quad \cdots \quad) &= \widehat{P}_2\\
f_3(\quad \cdots \quad) &= \widehat{P}_3\\
f_4(\quad \cdots \quad) &= \widehat{P}_4
\end{aligned}
$$

for $\widehat{\lambda}_{01}$, $\widehat{\lambda}_{02}$, $\widehat{\lambda}_{13}$, and $\widehat{\lambda}_{23}$. A simple iterative procedure for doing this was applied to data for both sexes and age groups from 15 to 85 years of age, using Model I. Thus far, no numerical work has been undertaken with Models II and III

Using the estimated parameters one can estimate the average probability of having an onset of mental disease in any one year between the ages of, for example, 35 to 39, the assumption being made that the underlying process is the same for all persons in this age group. For Model I this estimate is

$$
\text{EPO} = \frac{(\lambda_{01} + \lambda_{02})(1 - e^{-\lambda_{01}-\lambda_{02}-\lambda_{04}})}{(\lambda_{01} + \lambda_{02} + \lambda_{04})} .
$$

Since Model I tends not to reproduce the long tail of the lag time distribution, it therefore tends to under-estimate the proportion of persons having onsets who are never admitted to hospitals. A better estimate of the onset probability is very likely

$$\left[\frac{\text{EPO}}{\text{est. } P_{03}(\infty)}\right]\left[\frac{\text{Total hospitalized onsets with onset 35–39}}{\text{Total population of age 35–39}}\right]$$

where est. $P_{03}(\infty)$ is obtained by substituting the values of the estimated parameters into the expression for $P_{03}(\infty)$. Thus

$$\left[\frac{\text{EPO}}{\text{est } P_{03}(\infty)}\right]$$

can be looked upon as a correction factor to the first approximate estimate of the age specific onset rate (probability) obtained from dividing the total number of onsets of age x reaching a hospital by the total population of age x. Really one should not divide by the total population but include only those persons never admitted to a mental hospital, but this is a small correction except for the older age groups in the population.

The results of the next section of the paper indicate that the correction factor

$$\left[\frac{\text{EPO}}{\text{est } P_{03}(\infty)}\right]$$

is not very large, especially for the age groups less than 60 years of age. The reason is, of course, that the death rates are so low relative to the rates at which the insane enter hospitals that in only few cases death terminates life before admission can take place For the age groups over 60, however, the correction seems to run from 10 to 40%, i e., the factor ranges from 1 10 to 1.40 This result that the great majority of the insane eventually are hospitalized is not inconsistent with the results of prevalency studies which indicate that the number of insane found outside of institutions with no previous record of hospitalization is very large. Some rough calculations presented in the next section of the paper will show that there is no conflict in the case of our results. In any case it is clear that since the elapsed time between onset and admission is rather long in many cases, no contradiction in the results need exist. The prevalency consequences of our models should therefore be of considerable interest in themselves as well as forming the basis of an independent check on the general adequacy of the approximation the models offer of the onset-admission problems.

V. SOME RESULTS OF CALCULATIONS USING MODEL I

A. *Introduction*

In this section we wish to present in more detail some of the results obtained by calculations using Model I and our onset-admission data for Ohio state mental institutions for the years 1947–1948. Before turning to this, some comments about the spirit in which these calculations were made and the use of gross epidemiological models may serve to clarify our feelings about these calculations

In general it may be said that we are in favor of a lot of rough calculations, provided these imply some underlying model of the situation, in order to obtain from available data other numbers of interest. Our current work is, we believe, an example of this type of thing Psychiatric epidemiology requires, for most purposes, total incidence rates based on age of onset rather than hospital admission rates which are deficient both from the standpoint of incidence and age distribution We think that in the long run it is more important to attempt rough estimates of the rates in which we are really interested than to continue gathering quite exact data on admission rates. Proper caution must always be observed and something will be said of this below. Our feeling is that even models of the level of aggregation and grossness described earlier in this paper can produce interesting substantive results and at the same time serve to clarify the nature of the epidemiologic process

We do not, of course, believe that any model is as good as any other but we are quite sure that it will be some time before really good and also manageable models of the processes we are interested in will be produced From a certain practical point of view the reason one wishes to have a good model is that, more or less by definition, a good model is one which for one's particular purpose produces better—that is, more accurate—answers when used in calculations On this view good models are those which allow one to predict, with sufficient accuracy, interesting numbers There is another view which emphasizes or defines the goodness of models in terms of their power to unify lower order models, or in other words to combine in a nice way the fine structure of the process. Both functions of models are important; in this paper we are concerned primarily with the first.

Our attitude toward the calculations presented here is that they must be regarded as quite provisional Indeed, in the case of this particular problem, where several alternative models with no less *a priori* probability of giving a good fit to the data come easily to mind, our feeling is that until much further investigation of their predictive power

is undertaken, no results should be accepted which are not model independent. In other words, unless several or perhaps all reasonable models give approximately the same answer the result is not to be accepted.

We turn now to the discussion of the results using Model I.

B. *Onset-Admission Data, Ohio, 1947–1948*

The data which have made our studies possible have the characteristic that in addition to the usual age of admission information the duration of time between the initial onset of the disease and admission is also given. Two states now record such information: Illinois and Ohio We have primarily used the Ohio data and only the work done with them is reported here. There are some, not too easily described,

TABLE I

OHIO, STATE MENTAL PATIENTS, MALE, AGE OF ONSET OF FIRST ADMISSION CASES DISTRIBUTED BY TIME LAG BETWEEN ONSET AND ADMISSION*

Age of Onset	Time Lag Between Onset and Admission in Years						
	0–1	1–3	3–5	5–7	7–11	over 11	Total
15–19	74	28	17	8	7	26	160
20–24	98	48	25	11	14	35	231
25–29	104	50	21	23	18	34	250
30–34	152	62	31	21	20	21	307
35–39	170	53	27	17	16	22	305
40–44	194	68	15	13	19	26	335
45–49	177	55	16	11	17	16	292
50–54	148	45	16	15	15	12	251
55–59	161	52	11	12	12	9	257
60–64	162	82	35	13	9	7	309
65–69	180	73	27	16	17	5	318
70–74	120	82	33	12	10	4	261
75–79	112	60	34	11	7	0	224
80–84	71	32	22	3	0	0	128
85–89	27	11	2	0	1	0	41
90–94	4	2	0	1	0	0	7
95–99	3	0	0	0	0	0	3
Totals	1957	803	332	187	182	217	3678

* Data presented here are for the years 1947–1948 The cases represented are first admissions with psychosis, i e , with cases having diagnoses without psychosis and primary character disorders omitted Only those cases for which information existed concerning the lag between onset and admission could be used, these represent approximately 80–85% of the total first admissions, with psychosis, during these two years.

reasons for believing that they are slightly more accurate than the Illinois data, but we are not convinced that this is the case.

Although the information on the date or age of onset is obtained by asking for the dates of first onset and of current onset, the data punched on our IBM cards is in terms of the elapsed time between onset (first and current) and admission. We have segregated for study the first admissions cases with psychosis, and concerned ourselves with the elapsed time between first onset and first admission. The elapsed time is coded in such a way that one may distinguish time intervals of: one month for elapsed times of less than one year and intervals of one year for elapsed times greater than one year. For our purposes we had to calculate a patient's age at onset by using the elapsed time between his onset and admission dates and his age at admission. Due to the discrete nature of the elapsed time code some mistakes will be made in estimating age at onset for individual patients but the mistakes largely offset one another as far as the whole group of patients are concerned Tables I and II give the basic data classified in the manner appropriate for

TABLE II

OHIO, STATE MENTAL PATIENTS, FEMALE, AGE OF ONSET OF FIRST ADMISSION CASES DISTRIBUTED BY TIME LEG BETWEEN ONSET AND ADMISSION*

Age of Onset	Time Lag Between Onset and Admission in Years						
	0–1	1–3	3–5	5–7	7–11	over 11	Total
15–19	89	30	11	17	14	20	181
20–24	177	47	15	20	20	24	303
25–29	249	78	41	24	26	30	444
30–34	204	81	36	22	28	32	403
35–39	188	58	31	16	29	18	340
40–44	164	64	28	26	18	24	324
45–49	147	52	30	20	27	21	297
50–54	139	68	28	24	13	21	293
55–59	133	56	24	11	11	12	247
60–64	110	62	21	11	17	7	228
65–69	103	64	31	16	16	1	231
70–74	83	70	34	21	16	3	227
75–79	76	55	28	13	12	1	185
80–84	54	32	11	6	4	0	107
85–89	18	13	4	0	1	0	36
90–94	5	2	0	0	0	0	7
Totals	1939	832	373	247	252	214	3857

* See footnote to Table I.

use in the models described earlier. Table VI and IX, below, show the changes in age distribution of first admission patients when they are grouped according to age of onset instead of the age of admission. It is the function of the model calculations to refine these approximations and to include estimated numbers of cases who are never hospitalized.

There are at least two possible biases in the data which deserve mention since their existence would serve to qualify any of our results. By bias in the data we mean here an imperfection in the data such that when used in Model I biased estimates of the parameters will be obtained. Possible biases are:

1. The older a patient is when admitted to a hospital the more likely it is that information of almost every kind about him will be unavailable. This is often the case in data gathered upon admission to mental hospitals. If this were true with regard to age of onset information in the Ohio data the percentage of patients whose age of onset is unknown would be a monotonic increasing function of age of admission. The magnitude of the bias effect would depend upon the degree of the relation between age and lack of knowledge The bias in our estimates of λ_{01}, λ_{02}, λ_{13}, and λ_{23} would come about for the following reason There will be a systematic under-representation of persons with long elapsed times between onset and admission On the average, under-estimating the length of elapsed time leads to an under-estimate of the number of insane persons who are "lost" to the hospital system by virtue of intervening death This in its turn would lead to an under-estimation of the amount by which admission rates must be corrected to secure total incidence age of onset rates. Fortunately we shall see that the Ohio data do not have this defect

In other cases this bias could be roughly corrected if necessary by distributing the admissions of any given age, whose age of onset is unknown, among the possible ages of onset in a manner proportional to those for which the ages of onset are known. This was not done in our calculations; only the admissions with known age of onset were used because the foregoing type of bias was not present in the data.

TABLE III

PERCENTAGE OF CASES WITH AGE OF ONSET UNKNOWN BY AGE AT ADMISSION AND SEX

	10–19	20–29	30–39	40–49	50–59	60–69	70–79	80–89	90–99
Male	21	21	23	20	22	21	21	16	14
Female	20	16	16	19	19	16	19	20	14

TABLE IV

ESTIMATED AND ASSUMED VALUES OF THE PARAMETERS OF MODEL I FOR OHIO, MALE ONSET DATA, 1947–1948

Age of onset	Parameters for Model I						
	$\hat{\lambda}_{01}$	$\hat{\lambda}_{02}$	$\hat{\lambda}_{13}$	$\hat{\lambda}_{23}$	λ_{04}*	λ_{14}‡	λ_{24}†
15–19	000106	000114	260	13 928	00178	00890	.01780
20–29	000233	000167	212	12 630	00245	00868	01735
30–34	000315	000283	209	12 756	00321	00833	01665
35–39	000279	000341	219	16 228	00438	.01140	02280
40–44	000236	000438	255	7 341	00640	01670	03340
45–49	000196	000414	205	7 039	00961	02263	04525
50–54	000313	000330	213	29 437	01446	03075	.06150
55–59	—	—	Not computed		—	—	—
60–64	000676	000485	394	33 008	03117	05500	11000
65–69	000937	000753	281	30 035	04570	07810	.15620
70–74	001500	000683	324	11 125	07384	13475	26950
75–79	002490	00119	237	14 247	10392	16595	33190
80–84	002320	00142	379	28 805	15238	21100	42200
85–99	—	—	Not computed		—	—	—

* Death rates (sane) taken from life table, white males, 1939–41

† It has been assumed that the severe cases have death rates as high outside the mental hospitals as all patients do within the hospitals This assumption does not seem to be too bad *a priori* and since the severe cases appear to make the transition to the hospital so quickly, at least in our model, an error in setting this rate would not have too much effect upon our calculations Malzberg's relative mortality corrections were used, except for the age group 15–19 which Malzberg did not treat In this case it was assumed that the relative mortality was 10 times the normal rate B Malzberg "Life tables for patients with mental disease," *Mental Hygiene*, Vol 16, 1932 These corrections are almost identical with those which would have been derived from O Oedegaard, "Mortality in Norwegian mental hospitals 1926–1941," *Acta Genetica et Statistica Medica*, Vol II, 1951

‡ It is assumed that mild cases have death rates one-half as high as the severe cases when outside mental hospitals This is a correction in the right direction but is essentially arbitrary for there is no information available on this matter

2. We conjecture that patients, relatives and therefore clinical records tend to under-estimate the elapsed time between first onset and admission Insidiousness of onset and other difficulties of observation probably lead to a postdating of the onset period In addition people seem to forget or be mistaken in their remembrances of the earliest occurrences of onset symptoms The very large number of cases reported with elapsed time between first onset and admission of less than one month probably supports our view.[4] The effect of this is to bias the

[4] Swedish data published by Larsson and Sjögren [5] show that only 37 per cent of hospitalized cases had a lag between onset and admission of one year or less, whereas 52 per cent of our cases fall in this group Eighteen per cent of the Swedish cases had a lag of 10 years or more, whereas only 6 per cent of our cases show a lag over 11 years.

estimates obtained from Model I in the direction of under-estimating the corrected onset rates; i.e , estimates of λ_{01} and λ_{02} will be too small.

3. There is great uncertainty about the death rates of the insane outside mental institutions, especially so with regard to the period before first admission has taken place The error in our rough guesses of

TABLE V

ESTIMATED AND ASSUMED VALUES OF THE PARAMETERS OF MODEL I FOR OHIO, FEMALE ONSET DATA, 1947–1948

Age of onset	Parameters of Model I						
	$\hat{\lambda}_{01}$	$\hat{\lambda}_{02}$	$\hat{\lambda}_{13}$	$\hat{\lambda}_{23}$	λ_{04}*	λ_{14}‡	λ_{24}†
15–24	—	—	Not computed			—	—
25–29	000331	.000424	232	15 568	00201	00775	01550
30–34	000336	.000377	252	10 846	00249	00787	01573
35–39	000270	000366	265	17 494	00323	00980	01960
40–44	000450	000349	142	10 812	00446	01270	.02539
45–49	000362	000330	151	11 058	00643	01440	02880
50–54	000508	000371	144	6 708	00945	01805	03610
55–59	000305	000436	273	6 657	01421	02275	.04550
60–64	000391	000459	.267	4 666	02279	03465	06930
65–69	.000706	000480	229	5 976	03438	04505	09010
70–74	001462	.000461	199	8 066	05561	06900	13800
75–79	001916	000644	239	18 208	.08854	10850	.21700
80–84	002237	000904	240	21 767	13557	16103	32205
85–94	—	—		Not computed		—	—

* Death rate (sane) taken from life table, white females, 1939–41

† See note † to Table IV.

‡ See note ‡ to Table IV.

the death rates of the insane outside institutions is almost certainly more important than the structural bias of the Model that we have just mentioned

4 Apart from these data biases the Model itself has a structural bias This bias is the result of holding death rates constant when in fact they are increasing For example, in our calculations a man who has his first onset at age 50 and is first admitted at age 60 is treated as though he were a member of a cohort which over this ten year period was subject to death rates appropriate to ages 50–54 rather than to death rates which increase each year This again tends to produce under-estimates of λ_{01} and λ_{02}

C. *Results of Computation with Model I*

In Tables IV and V are given the results of our calculations using the Ohio data, 1947–1948, and Model I. The assumed death rates λ_{04}, λ_{14}, and λ_{24} are given as well as the estimated values of λ_{01}, λ_{02}, λ_{13}, and λ_{23} A few cases were not computed. This was because for those age groups the data had a character which made it inappropriate to the model and the technique of computation we were using. Thus among females, ages 15–19, there are 11 cases with 3–5 years elapsed times between onset and admission and 17 cases with 5–7 years elapsed times between onset and admission, whereas our method of calculation requires that the number of cases be a strictly decreasing function of elapsed time If this is not so inadmissible estimates of some of the parameters are obtained; i e , negative transition rates between states are obtained In other models this may not be required and with other computational methods certainly will not be required. For example, if we had used Neyman's minimum χ^2 methods with five or six elapsed time groups rather than the method of moments technique with four elapsed time groups this exclusion would not have been necessary.

In order to judge how variable from sample to sample the estimates of λ_{01}, λ_{02}, λ_{13}, and λ_{23} might be some rough estimates of their sampling variances have been made Rather than differentiating the expressions for $P_{03}(t)-P_{03}(t-j)$ with respect to $\lambda_{01}, \cdots, \lambda_{23}$ we obtained an approximation of the appropriate quantities by solving linear equation systems derived from the data. From age group to age group the $P_1(t), \cdots, P_4(t)$ change (t designating the age group) producing concomitant variation in the estimates $\hat{\lambda}_{01}, \cdots, \hat{\lambda}_{23}$ Solving the obvious equation systems one obtains, vector by vector, a rough estimate of

$$D = \begin{pmatrix} \frac{\partial\lambda_{01}}{\partial P_1} & \frac{\partial\lambda_{01}}{\partial P_2} & \frac{\partial\lambda_{01}}{\partial P_3} & \frac{\partial\lambda_{01}}{\partial P_4} \\ \frac{\partial\lambda_{02}}{\partial P_1} & \frac{\partial\lambda_{02}}{\partial P_2} & \frac{\partial\lambda_{02}}{\partial P_3} & \frac{\partial\lambda_{02}}{\partial P_4} \\ \frac{\partial\lambda_{13}}{\partial P_1} & \frac{\partial\lambda_{13}}{\partial P_2} & \frac{\partial\lambda_{13}}{\partial P_3} & \frac{\partial\lambda_{13}}{\partial P_4} \\ \frac{\partial\lambda_{23}}{\partial P_1} & \frac{\partial\lambda_{23}}{\partial P_2} & \frac{\partial\lambda_{23}}{\partial P_3} & \frac{\partial\lambda_{23}}{\partial P_4} \end{pmatrix}$$

the matrix over the range of ages considered. In our case the computations were performed using the data for males ages 30–34 to 50–54. The covariance matrix Σ of the vector of estimates $(\hat{\lambda}_{01}, \cdots, \hat{\lambda}_{23})$ is then

$$\Sigma = D\Omega D'$$

where Ω is the covariance matrix of $P_1(t)$, $P_2(t)$, $P_3(t)$, and $P_4(t)$. Because of the cross-sectional nature of our data the P's are not independent, being very slightly negatively correlated. Therefore considering them as independent will not give too bad an estimate. We find the approximate standard errors to be as follows:

Parameter	*Average value (30–34 to 50–54)*	*Estimate standard error*
λ_{01}	.000268	.000078
λ_{02}	.000361	.000045
λ_{13}	.220	.029
λ_{23}	14.560	17.46

Because the other parameters, i e , λ_{04}, λ_{14}, and λ_{24}, also were varying, the estimates of the standard errors must be taken as extremely rough. Nonetheless as far as the estimates of λ_{01} and λ_{02} are concerned the evidence is that they do not have very large variances relative to their absolute values. In addition we are interested in the estimate of $\lambda_{01}+\lambda_{02}$ rather than λ_{01} and λ_{02} separately. λ_{01} and λ_{02} have a negative covariance so that $\lambda_{01}+\lambda_{02}$ has an estimated standard error of .000045, which compared with the average value of $\lambda_{01}+\lambda_{02}$ of .000629 is a percentage error of 7.2 per cent, certainly a very reasonable level of accuracy

TABLE VI

Onset age group	Estimated number of cases, lag 7–11 years	Observed number of cases, lag 7–11 years
15–19	5.6	7
20–24	6 9	14
25–29	—	—
30–34	23 8	20
35–39	17 4	16
40–44	21 1	19
45–49	12.9	17
50–54	27 4	15
55–59	—	—
60–64	7.4	9
65–69	15 1	17
70–74	6.0	10
Mean value	12 80	14.40

In general the fit of the model to the data appears fair. As a test of goodness of fit we have used the data for the various age groups on the number of cases with a time lag of from 7 to 11 years between onset and admission. The data for the males give the results in Table VI. Since the standard error in each case is essentially equal to the square root of the expected number of cases, the discrepancies between expected and observed numbers of cases is close in most cases The value of χ^2 for the ten age groups is 18 752. The probability of exceeding this number with 10 degrees of freedom is approximately .04 In view of the known tendency of the model to underestimate the tail of the lag distribution the results are not discouraging. The fit to the female data is similar.

D. *Estimates of True Onset Rates Derived from Model I Calculations*

The main substantive result we have hoped to produce is a set of estimates of age specific (first) onset rates Even given the basic results of the calculations there are several alternative estimates one might produce From among these we have chosen as giving the best estimate the expression

$$\left[\frac{\text{EPO}}{\text{est. } P_{03}(\infty)}\right]\left[\frac{\text{Total hospitalized onsets of age } x}{\text{Total population of age } x}\right]$$

evaluated for each age group. By EPO we mean the estimated probability of having an onset at say age x (during the year in which the person is of age x) obtained by substituting in the appropriate expression for Model I

$$\frac{(\lambda_{01}+\lambda_{02})(1-e^{-\lambda_{01}-\lambda_{02}-\lambda_{04}})}{(\lambda_{01}+\lambda_{02}+\lambda_{04})}$$

the estimated and assumed values of the parameters By est. $P_{03}(\infty)$ we mean, similarly, the estimated probability of ever being admitted to a mental hospital based on the Model I results The ratio of these two estimates represents a correction factor with which to multiply the crude estimate of the onset probability at age x. The expression for this correction factor is

$$\frac{(\lambda_{01}+\lambda_{02})(\lambda_{13}+\lambda_{14})(\lambda_{23}+\lambda_{24})}{\lambda_{01}\lambda_{13}(\lambda_{23}+\lambda_{24})+\lambda_{02}\lambda_{23}(\lambda_{13}+\lambda_{14})}.$$

In Table VII are given the estimated values of this correction factor for the various age groups, by sex. For the ages below 65 not much of a

correction is involved but above this age it is substantial. This reflects principally the rapid increase in the death rates above this age since in general the average elapsed time between onset and admission tends to decline very slightly as age of onset increases.

Since readers will be more familiar with age specific first admission rates, we will derive our final correction factors so that they may be

TABLE VII

ESTIMATED RATIOS OF ONSET PROBABILITY TO ADMISSION PROBABILITY OF PERSONS HAVING ONSETS AT A GIVEN AGE, MODEL I, OHIO 1947–1948

Age group	Ratio of onset probability to admission probability	
	Male	Female
15–19	1.02	—
20–24	1.02	—
25–29		1.02
30–34	1.02	1.05
35–39	1.02	1.02
40–44	1.03	1.05
45–49	1.04	1.05
50–54	1.07	1.07
55–59	—	1.04
60–64	1.08	1.07
65–69	1.20	1.12
70–74	1.27	1.25
75–79	1.40	1.31
80–84	1.29	1.41

applied directly to admission rates. In Tables VIII and IX are given the ratios, by age group and sex, of the number of first admissions in a given age group to the number of first onsets in the same age group. Since Tables VIII and IX include only hospitalized patients the number of onsets is too small. By applying the correction factors in Table VII we can secure estimates of the total number of onsets in the population at large. The correction factors to be applied to age specific first admission rates in order to estimate total age specific onset rates can be obtained by dividing the entries in Table VII by the appropriate entry from Tables VIII and IX. The final results of this operation are given in Table X.

From Table X it appears that the very largest corrections are for the ages below 30. In the age group 15–19 hospital first admission rates

under-estimate the total onset rate by about 70 per cent and for the age group 25–29 by about 21 or 22 per cent. For the ages from 30 or 35 to 65 the average correction is small (6–7 per cent for males and 1 per cent for females) In this region first admission rates are evidently good estimates of the onset rate, assuming that the biases discussed earlier have not too great an effect. For the ages over 65 the picture is mixed. For a period of 10 years (65–75) in the case of females and 15 years (65–80) in the case of males the correction indicated is a 20 per cent

TABLE VIII

RATIO OF FIRST ADMISSIONS IN EACH AGE GROUP TO THE ONSETS IN EACH AGE GROUP, OHIO MENTAL PATIENTS (FIRST ADMISSIONS, WITH PSYCHOSIS), MALE, 1947–1948

Age groups	Number of onsets	Number of admissions	Ratio = Admissions / Onsets
15–19	160	96	60
20–24	231	164	.71
25–29	250	211	.84
30–34	307	276	.91
35–39	305	320	1 05
40–44	335	253	.75
45–49	292	316	1.07
50–54	251	262	1 04
55–59	257	283	1 10
60–64	309	298	.97
65–69	318	319	1 00
70–74	261	266	1.02
75–79	224	260	1.16
80–84	128	170	1.33
85–	51	82	1 61
Total	3678	3678	—

increase. After this period the correction becomes small as far as our estimates indicate, but they are very unstable for these ages.

For the very young age groups the large correction factor (to adjust admission rates as estimates of onset rates) is entirely the result of the shift to age of onset as the accounting variable. Admission rates are rising sharply between the ages of 15 and 30 and the shift to age of onset means a shift of many more cases into these age groups than out of them since the lag between onset and admission does not vary much

with age. Death rates are very low and few cases are lost between onset and admission. For the middle age group, 30 to 65, no correction is found because admission rates are relatively stable over this period and the reallocation of cases on the basis of age of onset merely replaces one case for another on the average. Death rates still do not have much effect. For the older age group, on the other hand, two factors produce

TABLE IX

RATIO OF FIRST ADMISSIONS IN EACH AGE GROUP TO THE ONSETS IN EACH AGE GROUP, OHIO MENTAL PATIENTS (FIRST ADMISSIONS, WITH PSYCHOSIS), FEMALE, 1947–1948

Age groups	Number of onsets	Number of admissions	Ratio $= \frac{\text{Admissions}}{\text{Onsets}}$
15–19	181	108	60
20–24	303	257	85
25–29	444	367	.83
30–34	403	397	.99
35–39	340	361	1.06
40–44	324	312	97
45–49	297	315	1.06
50–54	293	293	1 00
55–59	247	292	1 18
60–64	228	240	1 05
65–69	231	228	.99
70–74	227	233	1 03
75–79	185	232	1 25
80–84	107	142	1.33
85–	43	80	1.86
Total	3857	3857	—

the rather substantial correction factors: admission rates are again rapidly increasing and death rates are high with consequent high attrition between onset and admission

E *Some Comments on the Consequences of the Calculations for Prevalency Estimates Derived from Model I*

A matter of some interest, both as an independent check on the results of the computations and adequacy of the model and as an additional result of substantive interest, is the prediction, from our calculations, of one type of prevalency rate for the various age groups. By age-

specific prevalency rates we mean the total number insane on any given date in the population within given age groups. Mental hospital reports provide, of course, exact data on the number of hospitalized insane in the various age groups Our problem is to supplement these numbers with estimates of the number of insane not hospitalized. The insane outside of hospitals may be grouped into two classes: (*1*) Cases not previously hospitalized, (*2*) Cases previously hospitalized and now

TABLE X

TOTAL CORRECTION FACTORS FOR AGE-SPECIFIC FIRST ADMISSION (WITH PSYCHOSIS) RATES, OHIO, 1947–1948, TO OBTAIN AGE-SPECIFIC ONSET RATES

Age groups	Total correction factors for first admission	
	Male	Female
15–19	1 700	1.700*
20–24	1 436	1.200*
25–29	1 215	1 229
30–34	1 122	1.061
35–39	971	.962
40–44	1 373	1.082
45–49	972	.991
50–54	1 029	1 070
55–59	.881†	.881
60–64	1 114	1 019
65–69	1 200	1.131
70–74	1 245	1 214
75–79	1 207	1 048
80–84	970	1 060

* These female age groups are assumed to have the same entries in Table VII as the corresponding male age groups

† This male age group is assumed to have the same entry in Table VII as the corresponding female group

again insane In our models the not previously hospitalized group is represented by the members of the various unhospitalized insane states; e g., in Model I members of states S_1 and S_2, in Model II members of state S_2. None of our models has provision for discharged cases The actual calculation for Model I of $P_{01}(t)+P_{02}(t)$, the probability of being in either state S_1 or S_2 at time (age) t, would be quite tedious except for say the ages 15–19; even then one would have to assume that the onset rate was zero up until age 15 For other ages to do the calculations

correctly it would be necessary to piece together the contributions from all of the processes up to the age required Thus, if one wished to estimate the prevalency rate for cases not previously hospitalized at age 40, one would have to compute the contribution of the process during the ages 35–39 with its set of parameters, compute the contribution of the process during the ages 30–34 with its set of parameters, etc. Even this would only be an approximation due to the assumption of constant parameter values over time which has been mentioned earlier While all this is true, by the use of ruthless approximations something can be said about the number of cases one would expect to find in a prevalency survey that have not previously been treated in hospitals for the insane This does not solve the problem of computing total prevalency rates but does offer a check of our onset calculations and is a result of some direct interest Oedegaard has made some similar calculations with a simple model for Norwegian data [8].

If we collapse Model I into Model I′ and average the onset and lag parameters in the obvious way, we get the following approximations. First, notice that the admission rate at any time t is

$$\frac{dS_3(t)}{dt} = \frac{dP_{03}(t)}{dt} = \bar{\lambda}_{13}(t)S_1(t);$$

thus

$$S_1(t) \sim \frac{\text{First Admission Rate at } t}{\bar{\lambda}_{13}(t)},$$

where

$$\bar{\lambda}_{13}(t) = \left\{\frac{\lambda_{01}\left\{\dfrac{1}{\lambda_{13}}\right\} + \lambda_{02}\left\{\dfrac{1}{\lambda_{23}}\right\}}{\lambda_{01} + \lambda_{02}}\right\}^{-1}.$$

This will make the average elapsed time between onset and admission the same in the two processes: Model I and Model I′. In Table XI are given the values of $\bar{\lambda}_{13}$ for the various age groups, by sex Mean elapsed time between onset and admission, $\bar{\lambda}_{13}$, is equal to the ratio of the age specific first admission rate to the proportion of the population of given age, insane but not previously hospitalized. The values obtained, shown in Table XI, check rather well with some earlier calculations of ours based on World War II Selective Service data for the male age

group 18–34 where we estimated this ratio to be approximately 2.23. Oedegaard's calculations lead to an average value of $\bar{\lambda}_{13}$ on the order of 3.

In order to obtain what are ordinarily thought of as age-specific prevalency rates it would be necessary to compute from other types of data the remaining components: The proportion of persons in some age group that are residents of mental hospitals and the proportion of persons, insane, not in hospitals, and with a record of previous hospitalization. Our models are not of much help in obtaining estimates of either of these two components. An estimate of the first component is readily available from hospital records, but the second component would have to be estimated from discharge and readmission data, or perhaps from some prevalency survey of a follow-up, case study type. While our particular models do not contribute to the determination of the required estimates, models of the same type might prove very useful in the analysis of the discharge-readmission process.

TABLE XI

PREVALENCY IMPLICATIONS OF MODEL I′ WITH USE OF ESTIMATED PARAMETERS, OHIO, 1947–1948, PERSONS WITH PSYCHOSIS

Age group	Values of λ_{13} = Age specific first admission rate / Proportion of population of given age, insane but not hospitalized by specified age	
	Male	Female
15–19	1 88	—
20–24	2.78	—
25–29		2 47
30–34	2.57	1 92
35–39	2 09	2.00
40–44	1 46	3.99
45–49	2 15	4 15
50–54	2 32	4 05
55–59	—	1 59
60–64	1 49	1.86
65–69	1.97	2 64
70–74	2.15	3 84
75–79	2.69	3.15
80–84	1 65	2.97
Average $\bar{\lambda}_{13}$	2.10	2 89

APPENDIX

Analysis of Model I

In this Appendix we work out in more detail the probability expressions required for the application of one of the three models to admission data.

In order to achieve our ultimate goal in the analysis of onset and admission data we must evaluate the probability of an individual being in S_3 at the end of a time period consisting of n basic time intervals, given that the individual was in S_0 at the beginning of the period. Thus we will assume that for the particular age group under consideration all individuals are in state S_0 at the beginning of a certain time period. After n days (at the end of time period t_n), and taking one day as the basic time interval, we would like to know how many persons in this age group will be in mental hospitals. This is equal to the number of persons in the cohort multiplied by the probability of an individual of the cohort going from S_0 at the beginning of t, to S_3 by the end of t. It is the great virtue of the matrix formulation that it leads quite easily to the evaluation of such probabilities. Let us define $p_{ij}^{(n)}$ as being the probability of being in state j after n steps if one began in state i. The p_{ij}^{n} are the i, jth elements of the nth power of the matrix so that in Model I

$$p^n = \begin{pmatrix} P_{00}^{(n)} & P_{01}^{(n)} & P_{02}^{(n)} & P_{03}^{(n)} & P_{04}^{(n)} \\ 0 & P_{11}^{(n)} & 0 & P_{13}^{(n)} & P_{14}^{(n)} \\ 0 & 0 & P_{22}^{(n)} & P_{23}^{(n)} & P_{24}^{(n)} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Special techniques are available for the evaluation of the nth powers of matrices, especially the elements of P^n associated with absorbing states in rather simple matrices [1].

As indicated in the body of the text the process is broken into two parts characterized by the matrices

$$P = \begin{pmatrix} P_{00} & P_{01} & P_{02} & 0 & P_{04} \\ 0 & P_{11} & 0 & P_{13} & P_{14} \\ 0 & 0 & P_{22} & P_{23} & P_{24} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

and

$$P_{00} = \begin{Bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & P_{11} & 0 & P_{13} & P_{14} \\ 0 & 0 & P_{22} & P_{23} & P_{24} \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Bmatrix}.$$

We are to evaluate P^{365}, $P^{365}P_{00}{}^{365}$, $\cdots$; $P^{365}P_{00}{}^{365K}$, $\cdots$ Let us for the present replace 365 with a generalized N. As a first step we obtain $P_{01}{}^{(N)}$, $P_{02}{}^{(N)}$, and $P_{03}{}^{(N)}$, the latter expression first Define λ_n as the probability of being in S_3 for the first time at the nth step, $n \leqq N$. Then

$$\lambda_n = p_{01}{}^{(n-1)}p_{13} + p_{02}{}^{(n-1)}p_{23},$$

or, verbally, λ_n equals the probability of being in S_1 at the $n-1$th step and going from S_1 to S_3 at the nth step or the probability of being S_2 at the $n-1$th step and going from S_2 to S_3 at the nth step. Clearly the probability of being in S_3 at the Nth step is equal to $\sum_{n=1}^{N}\lambda_n$ since S_3 is an absorbing state We find that

$$p_{01}{}^{(n-1)} = \frac{(p_{11}{}^{n-1} - p_{00}{}^{n-1})p_{01}}{(p_{11} - p_{00})}$$

$$p_{02}{}^{(n-1)} = \frac{(p_{22}{}^{n-1} - p_{00}{}^{n-1})p_{02}}{(p_{22} - p_{00})}$$

and, thus

$$\lambda_n = \frac{(p_{11}{}^{n-1} - p_{00}{}^{n-1})p_{01}p_{13}}{(p_{11} - p_{00})} + \frac{(p_{22}{}^{n-1} - p_{00}{}^{n-1})p_{02}p_{23}}{(p_{22} - p_{00})}$$

$p_{03}{}^{(N)} = \sum_{n=1}^{N}\lambda_n$ and since these sums involve only geometric series, we have

$$\begin{aligned} p_{03}{}^{(N)} = {} & \frac{p_{01}p_{13}}{(p_{11} - p_{00})}\left[\frac{1 - p_{11}{}^N}{1 - p_{11}}\right] + \frac{p_{02}p_{23}}{(p_{22} - p_{00})}\left[\frac{1 - p_{22}{}^N}{1 - p_{22}}\right] \\ & - \left[\frac{p_{01}p_{13}}{(p_{11} - p_{00})} + \frac{p_{02}p_{23}}{(p_{22} - p_{00})}\right]\left[\frac{1 - p_{00}{}^N}{1 - p_{00}}\right]. \end{aligned} \tag{1}$$

It is convenient to replace this expression for the discrete process by its analog for a continuous process In order to do this we shrink the basic time intervals (of one day) down toward zero, keeping constant, at say

one year, the total length of the time period during which the process has been going on In order to do this we let $\lambda_{ij}=Np_{ij}$ for $i\neq j$ In the limiting procedure when $N\rightarrow\infty$, p_{ij} will approach zero since the λ_{ij} remain fixed Our limiting procedure implies the additional restrictions

$$p_{00} = 1 - p_{01} - p_{02} - p_{04}$$
$$p_{11} = 1 - p_{13} - p_{14}$$
$$p_{22} = 1 - p_{23} - p_{24}$$

and, in the limit, the probability of remaining in a given state approaches one. After substituting λ_{ij}/N for p_{ij} in (1), and taking the limit[5] we have

$$(2)\quad \begin{aligned} p_{03}(t = 1) = {} & \frac{\lambda_{01}\lambda_{13}[1 - e^{-\lambda_{13}-\lambda_{14}}]}{(\lambda_{01} + \lambda_{02} + \lambda_{04} - \lambda_{13} - \lambda_{14})(\lambda_{13} + \lambda_{14})} \\ & + \frac{\lambda_{02}\lambda_{23}[1 - e^{-\lambda_{23}-\lambda_{24}}]}{(\lambda_{01} + \lambda_{02} + \lambda_{04} - \lambda_{23} - \lambda_{24})(\lambda_{23} + \lambda_{24})} \\ & - \frac{\lambda_{01}\lambda_{13}[1 - e^{-\lambda_{01}-\lambda_{02}-\lambda_{04}}]}{(\lambda_{01} + \lambda_{02} + \lambda_{04} - \lambda_{13} - \lambda_{14})(\lambda_{01} + \lambda_{02} + \lambda_{04})} \\ & - \frac{\lambda_{02}\lambda_{23}[1 - e^{-\lambda_{01}-\lambda_{02}-\lambda_{04}}]}{(\lambda_{01} + \lambda_{02} + \lambda_{04} - \lambda_{23} - \lambda_{24})(\lambda_{01} + \lambda_{02} + \lambda_{04})}. \end{aligned}$$

This is the probability of being in S_3 at the end of the first year for a person in S_0 at the beginning of that year. If we were interested in dividing this one year period into more than one smaller period, a "t" should appear in the exponent of all the exponential terms.

We must now turn our attention to those cases who while having their onset at age x do not enter the hospital (or S_3) until ages $x+1$, $x+2$, etc. As before, we define λ_n to be the probability of being in S_3 for the first time at the nth step, this time $n\geqq N$. Thus,

$$\lambda_n = p_{01}{}^{(N)}p_{11}{}^{(n-N-1)}p_{13} + p_{02}{}^{(N)}p_{22}{}^{(n-N-1)}p_{23}.$$

We see that

$$p_{11}{}^{(n-N-1)} = p_{11}{}^{n-N-1},$$
$$p_{22}{}^{(n-N-1)} = p_{22}{}^{n-N-1}.$$

---

[5] $$\lim_{N\rightarrow\infty}\left(1-\frac{\lambda}{N}\right)^N = e^{-\lambda}.$$

Substituting and summing,

$$\sum_{n=N+1}^{N'} \lambda_n = p_{01}^{(N)} \frac{p_{13}(1 - p_{11}^{N'-N})}{1 - p_{11}} + p_{02}^{(N)} \frac{p_{23}(1 - p_{22}^{N'-N})}{1 - p_{22}};$$

thus,

$$p_{01}^{(N)} = \frac{(p_{11}^N - p_{00}^N)p_{01}}{p_{11} - p_{00}},$$

$$p_{02}^{(N)} = \frac{(p_{22}^N - p_{00}^N)p_{02}}{p_{22} - p_{00}}.$$

Substituting and taking limits as before, we have

$$p_{03}(t > 1 \mid s_1 \text{ or } s_2 \text{ at } t = 1)$$

$$= \frac{\lambda_{01}\lambda_{13}\left[1 - e^{-(\lambda_{13}+\lambda_{14})(t-1)}\right]\left[e^{-\lambda_{13}-\lambda_{14}} - e^{-\lambda_{01}-\lambda_{02}-\lambda_{04}}\right]}{(\lambda_{01} + \lambda_{02} + \lambda_{04} - \lambda_{13} - \lambda_{14})(\lambda_{13} + \lambda_{14})} \tag{3}$$

$$= \frac{\lambda_{02}\lambda_{23}\left[1 - e^{-(\lambda_{23}+\lambda_{24})(t-1)}\right]\left[e^{-\lambda_{23}-\lambda_{24}} - e^{-\lambda_{01}-\lambda_{02}-\lambda_{04}}\right]}{(\lambda_{01} + \lambda_{02} + \lambda_{04} - \lambda_{23} - \lambda_{24})(\lambda_{23} + \lambda_{24})},$$

where $t \geqq 1$. This represents only part of the probability of being in s_3 and the total cumulative probability of being in S_3 at $t \geqq 1$ is equal to

$$p_{03}(t = 1) + p_{03}(t \geqq 1 \mid S_1 \text{ or } S_2 \text{ at } t = 1).$$

Model I'

A simplified model of the onset-admission process can be obtained from Model I by removing one of the insanity states Let us remove S_2, then the process is characterized by the matrix of transition probabilities

$$P = \begin{pmatrix} p_{00} & p_{01} & 0 & p_{04} \\ 0 & p_{11} & p_{13} & p_{14} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Thus, since $\lambda_{02} = \lambda_{22} = \lambda_{23} = \lambda_{24} = 0$,

$$p_{03}(t = 1) = \frac{\lambda_{01}\lambda_{13}\left[1 - e^{-\lambda_{13}-\lambda_{14}}\right]}{(\lambda_{01} + \lambda_{04} - \lambda_{13} - \lambda_{14})(\lambda_{13} + \lambda_{14})} \tag{4}$$

$$- \frac{\lambda_{01}\lambda_{13}\left[1 - e^{-\lambda_{01}-\lambda_{04}}\right]}{(\lambda_{01} + \lambda_{04} - \lambda_{13} - \lambda_{14})(\lambda_{01} + \lambda_{04})},$$

and

$$p_{03}(t \geq 1 \mid s_1 \text{ at } t-1) = \frac{\lambda_{01}\lambda_{13}\left[1 - e^{-(\lambda_{13}+\lambda_{14})(t-1)}\right]\left[e^{-\lambda_{13}-\lambda_{14}} - e^{-\lambda_{01}-\lambda_{04}}\right]}{(\lambda_{01} + \lambda_{04} - \lambda_{13} - \lambda_{14})(\lambda_{13} + \lambda_{14})}$$

Similar results have been obtained for Models II and III but since no computations have been done using them it was not thought worth while to present them here. The expressions for $P_{03}(t)$, etc. are substantially more complicated, especially in the case of Model III.

REFERENCES

[1] Feller, W., *An Introduction to Probability Theory and Its Applications*, John Wiley and Sons, 1950, Chapters 15 and 16.

[2] Fremming, Kurt H., "The Expectation of Mental Infirmity in a Sample of the Danish Population," *Occasional Papers on Genetics*, No. 7, The Eugenics Society and Cassell and Co. Ltd. 1951.

[3] Hollingshead A. B , and Redlich, F. C., "Social Stratification and Psychiatric Disorders," *American Sociological Review*, 18 (April 1953), pp 163–69.

[4] Kemkan, Paul, Tietze, Christopher, and Cooper, Marcia, "Mental Hygiene Problems in an Urban District: Second Paper," *Mental Hygiene*, XXVI (1942), 100–19. (Reports main results of 1936 survey of East Baltimore.)

[5] Larsson, Tage, and Sjörgen, Torsten, *A Methodological, Psychiatric and Statistical Study of a Large Swedish Rural Population*, Copenhagen, 1954. (Acta Psychiatrica, Supplement 89).

[6] Neyman, J., *First Course in Probability and Statistics*, Henry Holt and Co., 1950, 69–93

[7] Neyman, J , "Outline of Statistical Treatment of the Problem of Diagnosis," *Public Health Reports*, Vol 62, No 40, 1947

[8] Oedegaard, O., "The Incidence of Mental Disease as Measured by Census Investigations Versus Admission Statistics," *Psychiatric Quarterly*, 26, April 1952, 212–218

[9] Roth, W. F., and Luton, F H. "The Mental Health Program in Tennessee," *American Journal of Psychiatry*, 99 (1943), 662–675 (Reports on 1938 survey of Williamson County, Tennessee)

MAXIMUM LIKELIHOOD AND MINIMUM χ^2 ESTIMATES OF THE LOGISTIC FUNCTION

JOSEPH BERKSON
*Mayo Clinic and Mayo Foundation**

IN MANY situations commonly encountered in statistical practice, the minimum χ^2 and the maximum likelihood estimates are identical An example is the estimation of $P=1-Q$, the proportion of a population possessing some specified characteristic; this may be presented in a simple way:

The Pearson χ^2 is given by[1]

$$\chi^2 = \sum \frac{(o-e)^2}{e}, \tag{1}$$

where o is the observed frequency and e is the expected frequency.

In a random sample of N from the population, the expected frequency of individuals having the characteristic is PN, and the expected frequency of individuals not having it is QN If we observe r "haves" and $s=N-r$ "haven'ts," and define an observed $p=r/N$, and an observed $q=1-p=s/N$, the χ^2 table is

	Have	Haven't	Total
Observed	pN	qN	N
Expected	PN	QN	N

$$\chi^2 = \frac{(pN-PN)^2}{PN} + \frac{(qN-QN)^2}{QN} = \frac{N}{PQ}(p-P)^2. \tag{2}$$

If we estimate P on the principle that we shall take that value which minimizes the χ^2 of (2) and designate the estimate as $\hat{p}$ we obtain

$$\hat{p} = p = r/N \tag{3}$$

The estimate (3) is well known to be also the maximum likelihood estimate of P, if p is assumed to be binomially distributed around the true P.[2]

As in the case with the estimate of the binomial parameter, the mini-

* The Mayo Foundation is a part of the Graduate School of the University of Minnesota.

[1] Unless otherwise specified, the χ^2 referred to is the classic χ^2 of Pearson Appendix Note 1 should be read for the definition of the minimum χ^2 estimate.

[2] It is worth noting that in deriving the χ^2 estimate, we needed only to be able to write the expectation PN, but to derive the maximum likelihood estimate, we needed to assume that the observed p is

mum χ^2 estimate of the Poisson parameter λ from a sample of the distribution

$$P\{x_i\} = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \tag{4}$$

can easily be shown to be $\bar{x}$, the mean of the sample, and this is also the maximum likelihood estimate

There are, however, some situations, which also occur not infrequently in practice, in which the minimum χ^2 and maximum likelihood estimates are not identical Even in these situations, as the number in each sample increases indefinitely, the sampling distribution of both estimates becomes more nearly the same. Both estimates are "efficient" in the sense of Fisher [16], as well as "best asymptotically normal" in the sense of Neyman [20] These characteristics, however, refer to asymptotic properties, which are in the realm of what Fisher calls the "theory of large samples," where, to quote his words, "nothing that we say shall be true, except in the limit when the sample is indefinitely increased; a limit, obviously, never attained in practice [13]."

For finite samples, which is to say for any real statistical samples, small or large, the estimates may differ in their distributions, and the question arises, "Which is the better estimate?" Little or nothing is reliably known which will provide an answer to this question, although conjectural opinions in favor of the maximum likelihood estimate are frequently expressed For the last several years, I have been trying to accumulate some information, intended to help to clarify the problem, by calculations and experiments with actual samples, referring to situations which I have encountered in practice The present article reports the results of one series of such experiments

The parameters to be estimated are regression coefficients, or perhaps it is better to say the coefficients in a functional equation, and the equation concerned is the logistic function with binomial variation of the dependent variate, which has been advanced [1] as a model for bio-assay with quantal response,

$$P_i = 1 - Q_i = \frac{1}{1 + e^{-(\alpha+\beta x_i)}}; \quad \hat{p}_i = 1 - \hat{q}_i = \frac{1}{1 + e^{-(a+bx_i)}} \cdot \tag{5}$$

distributed binomially in random samples It is an important methodological advantage of least-squares types of estimates, over the maximum likelihood estimate, that the calculation of the maximum likelihood estimate requires complete knowledge of the distribution, in order to enable one to write the specific probability which is to be maximised, whereas the method of least squares generally requires only some limited knowledge about some functions of the distribution, perhaps only the mean, perhaps the mean and variance.

The straight line transform of this function, called the logit, is given by

$$\text{logit } P_i = \ln P_i/Q_i = \alpha + \beta x_i. \tag{6}$$

The equations to be solved for the maximum likelihood estimate of α and β are

$$\sum n_i(p_i - \hat{p}_i) = 0 \tag{7}$$

$$\sum n_i x_i(p_i - \hat{p}_i) = 0, \tag{8}$$

where n_i is the number at x_i, $p_i = 1 - q_i$ is the proportion of n_i observed to respond, and $\hat{p}_i$ is the estimate of P_i

For the minimum χ^2 estimate the equations to be solved are

$$\sum n_i \frac{(\hat{p}_i q_i + \hat{q}_i p_i)}{\hat{p}_i \hat{q}_i} (p_i - \hat{p}_i) = 0, \tag{9}$$

$$\sum n_i \frac{(\hat{p}_i q_i + \hat{q}_i p_i)}{\hat{p}_i \hat{q}_i} x_i(p_i - \hat{p}_i) = 0. \tag{10}$$

Since the coefficients of these equations are functions of the $\hat{p}$ which is to be estimated, as well as because $\hat{p}$ is not linear in the parameters, in general, although not always, both the minimum χ^2 and maximum likelihood estimates require iterative methods for solution.[3] For situations met routinely in the laboratory, the iterative procedure is unsatisfactory, not only because of the large amount of arithmetic labor involved, but because, since the solution is only approached as a limit which is never really reached, there is a problem of defining the estimate which actually is attained [1].

In 1944 I [4] suggested a noniterative solution which I then designated as "least squares," but which is presently called the "minimum logit χ^2 estimate," defined by minimization of the following quantity, called the "logit χ^2"

$$\chi^2 \text{ (logit)} = \sum n_i p_i q_i (l_i - \hat{l}_i)^2 \tag{11}$$

where $l_i = \ln p_i/q_i$ represents the observed logit, and $\hat{l}_i = a + bx_i$ represents the estimated value of the logit The normal equations for obtaining the minimum logit χ^2 estimate of α and β are

$$\sum n_i p_i q_i (l_i - \hat{l}_i) = 0, \tag{12}$$

$$\sum n_i p_i q_i x_i (l_i - \hat{l}_i) = 0. \tag{13}$$

[3] See Appendix Note 2 for methods used in sampling and calculation.

The evaluation of (12) (13) leads to a procedure that amounts simply to obtaining a least-squares solution of the straight line

$$\hat{l} = a + bx_i, \tag{14}$$

using $n_i p_i q_i$ as weight of the observation l_i. Since the coefficients are in terms of the known observations, and $\hat{l}$ is linear in the parameters, (12) (13) can be solved directly and simply.

The quantity on the right hand side of (11), which is the logit χ^2, is asymptotically distributed as χ^2, as is the classic Pearson χ^2, and for small numbers is very close to the Pearson χ^2 [3]. Taylor [28] has presented the proof that the minimum logit χ^2 estimate falls in the class of estimates R.B A N. of Neyman—"regular best asymptotically normal"—and that it therefore has the same asymptotic properties as the maximum likelihood estimate.

We may now turn to an inquiry into the properties of these estimates for finite samples.

The criteria for judging which of several estimates is "best" are of course conceivably many One criterion frequently used is the comparative size of the mean square error, that is, the expected value of the squared difference of the estimate from the true value of the parameter; this is the one used in the present investigation.

The definitive part of the investigation was made for an experiment with three values of x_i equally spaced, unit distance apart and with $n_i = 10$ for each x_i. The true P's at the three successive positions of x_i were initially taken as 0 3, 0 5, 0 7, which defined the values of the parameters as $\alpha = 0$, $\beta = 0$ 84730. The program of experiments was in two sections, one in which β was considered known and only α to be estimated, the other in which α and β were to be estimated simultaneously, these two situations simulating actual practical conditions in the field of bio-assay.

With 10 animals exposed at x_i, there are 11 possibilities of number of animals responding, and with 3 doses there are in all 1,331 possible samples For the case with β known, α to be estimated, the statistics were calculated for the entire sampling distribution; that is, the estimates were evaluated for all 1,331 possible samples, and these were weighted by their appropriate probabilities to yield the required statistics. For the case with both α and β to be estimated, the total sampling distribution was evaluated in the case of the maximum likelihood and minimum logit χ^2 estimates, but because of the laborious calculations involved with the minimum Pearson χ^2 estimate, a stratified random

sample of 1,000 was used, for each of the dose arrangements for which experiments were performed [4]

MAIN EXPERIMENTS; β KNOWN, α TO BE ESTIMATED

For the first experiment the value of P corresponding to the central one of the 3 doses was 0.5 This experiment, disposed symmetrically around $P=0\ 5$, can reasonably be considered the model experiment, since in the design of a bio-assay experiment, one centers the dosages as nearly as possible at the E D. 50, the sampling errors of the estimates being smallest with this arrangement. Additional experiments were performed with central P at 0.6, 0.7, 0 8, and 0.85 The results are summarized in Table 1.[5] It is seen that at central $P=0.5$ all three estimates are unbiased and that the variance is smallest for the minimum

TABLE 1

ESTIMATE OF α, β KNOWN

True P at dose			Mean			Variance			Mean square error		
Low	Mid	High	Max lik.	Min Pearson χ^2	Min logit χ^2	Max lik	Min Pearson χ^2	Min logit χ^2	Max lik	Min Pearson χ^2	Min logit χ^2
3	5	7	0	0	0	158	139	137	158	139	137
.391	6	778	+ 012	− 017	− 026	164	143	141	164	144	141
5	7	845	+ 028	− 036	− 055	186	161	156	187	162	159
632	8	903	+ 056	− 059	− 097	246	207	187	249	.211	196
708	85	930	+ 080	− 077	− 141	305	249	201	312	.255	221

Logistic function $\beta=0\ 84730$ known, $\alpha=0$ to be estimated 3 equally spaced doses, 10 at each dose Comparison of statistics of the three estimators for various positions of the dosages. Based on total sampling population The two samples with observed p's respectively 0, 0, 0, and 100, 100, 100 per cent, yield an infinite estimate by maximum likelihood and were omitted in calculation of all the estimates The fraction of the total population constituted by such samples is very small, 0 3 per cent for the experiment with central $P=0\ 85$ and less than that for the other experiments

logit χ^2 estimate, next larger for the minimum Pearson χ^2 estimate, and largest for the maximum likelihood estimate. For all other dose arrangements each of the estimates is biased, while the mean square error, and also the variance about the mean, are again smallest for the minimum logit χ^2 estimate, largest for the maximum likelihood estimate, with the minimum Pearson χ^2 estimate falling between the two

MAIN EXPERIMENTS; α AND β BOTH TO BE ESTIMATED

For comparison of the mean square error of the estimates, experiments were performed with central P at 0 5, 0 6, 0.7, 0 8, but were not

[4] See Appendix Note 2 for methods used in sampling and calculation

[5] Results are shown for central $P\geq 0\ 5$ For experiments with central $P<0\ 5$, the results are the same as for the symmetrically placed experiment in which $P>0.5$, with only the modification that there is a change of sign of bias where there is bias.

done with central P beyond $P=0.8$, because with the two parameters to be estimated samples insoluble by maximum likelihood become too frequent.[6] The comparisons are shown in Table 2. The characteristics of the findings are worth noting in a few details.

For the estimate of α, with dosages disposed symmetrically around $P=0.5$, the estimates are unbiased, as they are in the case where only α is to be estimated [7] For all other dose arrangements the estimates are biased even as was seen to be the case with only α to be estimated. However, whereas in the case of α only to be estimated, the bias of

TABLE 2

True P at dose			Mean			Variance			Mean square error			Per cent samples insoluble by max. lik.
Low	Mid	High	Max. lik.	Min Pearson χ^2	Min logit χ^2	Max lik	Min Pearson χ^2	Min. logit χ^2	Max lik	Min. Pearson χ^2	Min logit χ^2	
ESTIMATE OF α												
3	5	7	0	.002	0	187	179	154	187	179	154	0 07
391	6	778	− 006	− 013	− 020	230	.218	.205	230	218	206	0 11
5	7	845	− 021	− 011	− 013	430	412	393	430	412	394	0 54
632	8	903	− 026	037	084	1 102	970	682	1 103	972	689	3 91
ESTIMATE OF β												
.3	.5	7	095	0624	048	313	276	268	322	.280	271	
.391	6	778	100	0620	.038	331	303	270	341	307	272	
.5	.7	.845	108	037	004	.393	322	274	.404	323	274	
.632	.8	.903	088	− 019	− 077	.458	392	.202	466	.392	208	

Same as Table 1, but α and β both to be estimated Comparison of statistics of the three estimators for various positions of the dosages Statistics of the maximum likelihood and minimum logit χ^2 estimates based on total sampling population, those of minimum Pearson χ^2 on stratified random sample of 1000 at each dosage arrangement. Samples not yielding finite estimates by maximum likelihood omitted in calculating all statistics

the maximum likelihood estimate was positive for all dispositions of the dosages, it is in the present situation negative, and increases in absolute value as the difference of central dose from $P=0$ 5 increases For both minimum χ^2 estimates the bias is negative in the experiment with central $P=0.6$, but this negative bias decreases in absolute value as the central P of the experiment is removed further from $P=0.5$, and is positive at central $P=0.8$ At some point between $P=0$ 7 and $P=0$ 8, we may surmise that the minimum χ^2 estimates of α are unbiased Since the negative bias increases for the maximum likelihood

[6] See Appendix Note 3 for a discussion of samples insoluble by maximum likelihood

[7] The bias of 0 002 shown for the minimum Pearson χ^2 estimate is a sampling error

estimate continuously with increase of α (change of central P of dosage arrangement), there is a zone of α for which the bias of the minimum χ^2 estimates is *less* than that of the maximum likelihood estimate, a situation which does not obtain when α alone is to be estimated. At central $P=0.7$, the bias of all three estimates is negative and that of the χ^2 estimates is smaller than that of the maximum likelihood estimates. As respects the variance and mean square errors, for all dosage arrangements of the experiments these are smallest for the minimum logit χ^2 estimates, next larger for the minimum Pearson χ^2 estimates, and largest for the maximum likelihood estimate, as was the case with only α to be estimated

As regards the estimate of β, at disposition of dosages central $P=0.5$, the bias situation is different from what was found for estimate of α, either with β known or to be estimated simultaneously with β. Here all three estimates are biased, the maximum likelihood showing the largest bias, the minimum Pearson χ^2 the next smaller, and the minimum logit χ^2 the smallest bias. It is to be recalled that in the case of the one parameter α to be estimated, the mean square errors of the minimum χ^2 estimates were smaller than those of the maximum likelihood estimates in spite of a larger bias, because of the smaller variance. Here, although the relations of the biases are different with different disposition of dosages, the bias as well as the variance, in a zone of dosage arrangements, is less for the χ^2 estimates than for the maximum likelihood estimate As regards variance around the mean, and mean square error, the comparisons are uniformly favorable for the χ^2 estimates at all dose arrangements As before for α to be estimated with β known or unknown, both the variance and mean square error of the estimate of β are smallest for the minimum logit χ^2 estimate, next larger for the minimum Pearson χ^2 estimate, and largest for the maximum likelihood estimate

A reader of the manuscript of this article suggested that it is worth emphasizing that the biases of the estimates are not large. To this may be added the observation that the biases of the minimum logit χ^2 estimate and that of the maximum likelihood estimate, such as they are, are practically equal (though of opposite sign) and that the minimum logit χ^2 estimate nowhere achieves its smaller mean square error from a smaller bias alone. For centrally placed dosages, which is the arrangement to which a well-designed experiment approximates, for estimates of α the bias is zero. Therefore, although we are not dealing with a system of unbiased estimates, we can from a practical view-

point fairly disregard the question of bias and consider the comparisons as essentially in terms of the variances.

SUPPLEMENTARY EXPERIMENTS

The experiments which have been described are the basic ones of the present investigation and involve calculations of the total sampling population. Supplementary to these I did some experiments, having in mind specific questions that arose in the course of a discussion of the results described in previous sections of this paper. These questions had to do mostly with whether the spread of the dosages and/or their number might reverse the comparison of results as between the maximum likelihood and minimum χ^2 estimates Three experiments were performed, comparing maximum likelihood and minimum logit χ^2 estimates, employing in each 100 stratified samples, equally spaced

TABLE 3

ESTIMATE OF α AND β: DOSAGES SPREAD

Statistic	a		b	
	Max. lik.	Min. logit χ^2	Max lik.	Min logit χ^2
Mean	035	034	.087	−.181
Variance	.287	186	352	.182
Mean square error	288	.187	.360	.215

Three equally spaced doses, with respective $P=0$ 1, 0.5, 0 9 10 at each dose $\alpha=0$, $\beta=2$ 197, both to be estimated Comparison of minimum logit χ^2 and maximum likelihood estimates Based on 100 stratified random samples, 13 per cent of samples with infinite estimate by maximum likelihood omitted in calculating all statistics

doses, 10 at each dose:[8] (1) 3 doses, central dose corresponding to $P=0.5$, lowest dose corresponding to $P=0.1$, highest dose to $P=0$ 9, both parameters to be estimated, in which situation about 13 per cent of the samples have an infinite estimate by maximum likelihood; (2) 11 doses equally spaced, central dose corresponding to $P=0$ 5, lowest dose corresponding to $P=0.01$, highest dose to $P=0$ 99, β known, α to be estimated; (3) 4-point parallel assay, lower dose corresponding to $P=0$ 3, upper dose to $P=0$ 7, 10 at each dose for both unknown and standard. The results, shown in Tables 3, 4, and 5, are entirely in keeping with those found in the main experiments Judged on the basis of variance

[8] See Appendix Note 2 on methods of sampling and calculation

about the mean or mean square error the minimum logit χ^2 estimate appears better than the maximum likelihood estimate.

Finally there should be mentioned the remarkable experiments of Taylor [27], though they are not a part of the present investigation. He explored the extreme situation of many doses up to 100, with only

TABLE 4

ESTIMATE OF α: 11 DOSES

Statistic	Max. lik.	Min. logit χ^2
Mean	.000	−.005
Variance	.103	.069
Mean square error	.103	069

Experiment with 11 doses equally spaced, 10 at each dose, $\beta=0$ 91902 known, $\alpha=0$ to be estimated. Lowest dose at $P=0$ 01, highest dose at $P=0$ 99 Comparison of minimum logit χ^2 and maximum likelihood estimates Based on 100 stratified random samples

TABLE 5

ESTIMATE OF LOG RELATIVE POTENCY

Estimator	Mean	Variance	M.S.E.
Max. lik	−.014	.197	.197
Min. logit χ^2	− 015	183	183

Four-point parallel assay, 10 at each of the four observations, lower dose $P=0.3$, upper dose $P=0.7$ Standard and "unknown" equal potency, relative potency $\rho=1$, $M=\log \rho=0$ Based on 100 stratified random samples

one at each dose, comparing the maximum likelihood estimate with the minimum Pearson χ^2 estimate His results are in agreement with those of the present study, that is, the mean square error of the maximum likelihood estimate was found to be larger than that of the minimum χ^2 estimate, the contrast being even more pronounced with one at each dose than in the present series which was standardized with 10 animals at each dose The mean square error of the maximum likelihood estimate, using 20 animals, was attained with the minimum χ^2 estimate using only 5 animals

THE ESTIMATES AND THE INFORMATION LIMIT OF THE VARIANCES

It was observed in the early phases of the present investigation, when only some results for the maximum likelihood and minimum Pearson χ^2 estimates were available, in the situation of β known, α to be estimated, that not only was the mean square error of the χ^2 estimate less than that of the maximum likelihood estimate, but that

the variance about the mean, as well as the mean square error, was less than $1/I$, where $I = E\ (\partial \ln \phi/\partial\alpha)^2$ is the "amount of information" of Fisher, ϕ being the probability of the total sample set. Since the quantity I, related to the reciprocal of the variance, is sometimes spoken of as the total amount of available information, the minimum Pearson χ^2 estimate seemed to be extracting more than the total amount of information present in the data! The quantity $1/I$ is the Cramér-Rao lower bound for the variance of a regular unbiased estimate, and since with P of central dose at 0 5, the minimum χ^2 estimate is unbiased, the finding that the variance was smaller than $1/I$ was in apparent contradiction with this fundamental law also At first it was speculated that the explanation of the paradoxical findings lay in the fact that there is always at least a small fraction of the total population of samples for which there is no finite solution either by minimum χ^2 or maximum likelihood, and that the Cramér-Rao theorem applies only when all the samples have finite estimates But a short time later, calculations were completed which provided the values of the minimum logit χ^2 estimates, this estimator as defined giving a finite estimate for all samples The variance was found to be even further below the Cramér-Rao bound than with the minimum χ^2 estimate!

These findings had the effect of upsetting an apple cart, for they seemed to cast doubt on the calculations which had been made and so on the entire investigation. The inquiry was in fact interrupted for a period of several months while a search was made for the error which I presumed had been committed Although the difficulty was eventually resolved, I am reporting the incident of the appearance of the dilemma, because it is instructive in clarifying some prevalent misunderstandings regarding the lower bound of the variance Also it affords me the opportunity of expressing great gratitude to Professor J. Neyman, who first put his finger on the exact point which was the source of the difficulty.

The complete formula for the lower bound of the mean square error of an estimate $\hat{\theta}$ of a parameter θ is given by the following inequality (assuming certain conditions of regularity).

$$E(\hat{\theta} - \theta)^2 \geq \frac{\left(1 + \dfrac{\partial b}{\partial \theta}\right)^2}{I} + b^2, \tag{15}$$

where $b = E(\hat{\theta} - \theta)$ is the bias of the estimate $\hat{\theta}$ [22] [9]

[9] This inequality is given incompletely in Cramér's [10] text, which was my source for the formula at the time referred to, and this was in part a reason for the confusion Professor Joseph Hodges first pointed out the error to me and gave me the correct formula

If $b=0$, then the mean square error becomes the variance, and it appears that its lower bound is given by the reciprocal of I, as stated previously. The question then was how to account for the finding that with the dosages placed symmetrically about central dose corresponding to $P=0.5$, the χ^2 estimates were unbiased, while their variances were less than $1/I$. The answer was given to me by Professor Neyman. He pointed out that even if $b=0$, if $\partial b/\partial\alpha\neq 0$, and has in fact a negative value, then the right hand side of (15) can be less than $1/I$. Such a situation can obtain if the estimate is unbiased for some but not all values of α. At the time that this suggestion was made, the values of b had been computed for a number of values of α,[10] but they had been recorded in the calculation sheets and had not been subjected to any special scrutiny The values of b were now recovered from the archives and the bias in relation to α was studied by graphic methods. The

TABLE 6

ESTIMATE OF α, β KNOWN: LOWER BOUND

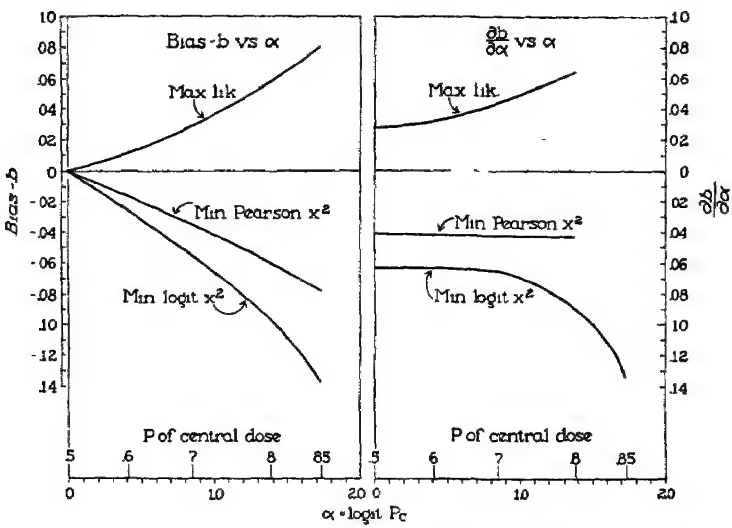
True P at dose			$1/I$	Maximum likelihood		Min Pearson χ^2		Min. logit χ^2	
Low	Mid	High		L B	M.S.E.	L B.	M S.E.	L.B.	M.S.E.
.3	.5	.7	149	158	158	137	139	.131	.137
.391	6	778	154	.164	164	142	.144	.135	.141
.5	.7	.845	169	185	.187	156	162	151	159
.632	8	903	208	.239	249	194	211	.182	.196

$\beta=0$ 84730 known $\alpha=0$ to be estimated, conditions as in Table 1. Values of $1/I=1/\Sigma n_iP_iQ_i$, and comparison of mean square error with lower bound.

bias function is shown in the figure. It is seen that for both the minimum Pearson χ^2 estimate and the minimum logit χ^2 estimate, the bias is either zero or negative, and its first derivative is indeed everywhere negative, while the bias for the maximum likelihood estimate is either zero or positive and its first derivative is positive. When the values of $\partial b/\partial\alpha$, estimated from the bias function graphically, and the computed value of b, were inserted in (15) for the correct calculation of the lower bounds, the mean-square errors of the χ^2 estimates fell properly above their respective lower bounds. These results are presented numerically in Table 6.

[10] A change of disposition of the dosages with samples considered as from a given logistic function such as one defined by $\alpha=0$, $\beta=0.84730$, is equivalent to samples drawn with the centrally disposed dosages unchanged but from functions with different values of $\alpha=\text{logit } P_c$, where P_c is the value of P corresponding to the central dose

Several points are to be made in connection with the findings shown in this table. *1.* The variance (Table 1) of the maximum likelihood estimate is larger and that of the χ^2 estimates is smaller than $1/I$. *2.* The lower bounds, the computation of which depends on the previously calculated values of the bias b, but not on the calculated values of the variances, are in the same order as found for the variances and the mean square errors; that is, the bound for the minimum logit χ^2 estimate is lowest, that for the minimum Pearson χ^2 estimate is next higher, and that of the maximum likelihood estimate is highest. This,



Relation of bias b, and $\partial b/\partial \alpha$ to α, shown for positive values of α For negative values of α, the bias reverses sign, while the sign of $\partial b/\partial \alpha$ remains the same.

it seems to me, can be considered as some confirmation of the basic validity of the findings respecting the order of the independently calculated variances and the mean square errors *3* In the case of the maximum likelihood estimate, at dosage arrangements with central $P=0\ 5$ and $P=0\ 6$, the mean-square errors and lower bound values are the same, to the precision of 3 significant figures as calculated; at central $P=0.7$, the difference between them is small, and only at central $P=0.8$ is the difference appreciable. Beyond central $P=0.8$—at $P=0.85$ for instance—the statistics were studied and the difference

between the mean square error and the lower bound was found to be even greater than at central $P=0.8$. The failure of the maximum likelihood estimate to attain its lower bound for the mean square error is perhaps surprising from general considerations of the properties of the maximum likelihood estimate, but it is in keeping with the specific findings of the present investigation. It is known that the lower bound is attained by an estimate only if the estimate is perfectly correlated with the logarithmic derivative $\partial \ln \phi/\partial\theta$. As is pointed out later (See Appendix note 4), the maximum likelihood estimator yields infinite estimates which are not in one-to-one relation with the logarithmic derivative, and therefore it could be anticipated that the maximum likelihood estimate cannot everywhere attain the lower bound value for its mean square error.

The logistic function with binomial variation of the dependent variate has sufficient statistics, which are $\sum n_i p_i$ and $\sum n_i x_i p_i$ for the estimate of α and β respectively.[11] Since sufficient statistics exist here, the possibility of improvement of the estimates by way of the Rao-Blackwell [24] [6] theorem presents itself. According to this remarkable theorem, if there is a sufficient statistic u for a parameter θ, and an estimate t which is not a one-to-one function of u, the conditional expectation of t given u is another estimate which has the same expectation as t and smaller variance than t. The maximum likelihood estimate, when it is finite, is necessarily a one-to-one function of the sufficient statistics $\sum n_i p_i$, $\sum n_i x_i p_i$, as can be seen directly from the equations of estimate, and the conditional expectation of the estimates given the sufficient statistic is therefore identical with the estimate itself; the maximum likelihood estimate is therefore not subject to improvement by means of this theorem. However, each of the minimum χ^2 estimates is subject to improvement. For the estimate of α, given β as known, the Rao-Blackwellized χ^2 estimates were computed directly, and their mean square errors were calculated. These are shown in Table 7. Here something extraordinary is evident—*within the precision of the arithmetic calculations, the mean-square errors of the minimum logit χ^2 estimates are equal to the previously and independently computed lower bounds, for all values of central P*. Thus this estimate attains what Rao [23] calls the information limit of the variance. I believe that the Rao-Blackwellized minimum logit χ^2 estimate is the first estimate achieved that has full efficiency in finite samples and has variance less than $1/I$. It is, of course, also sufficient.

---

[11] See Appendix 4 on the sufficiency of the estimates.

APPENDIX NOTE 1. THE DEFINITION OF THE MINIMUM χ^2 ESTIMATE

Some writers [12] [14] [18] have compared the minimum χ^2 estimate unfavorably with the maximum likelihood estimate, and their examples include the estimate of the binomial parameter P. Yet, as developed in the present paper, these two estimates are identical, so that one can hardly be inferior to the other. An examination of the works referred to discloses that there is ambiguity in respect to what is meant by the minimum χ^2 estimate. The contradiction brings to the

TABLE 7

ESTIMATE OF α, β KNOWN: RAO-BLACKWELLIZATION

P of central dose	Minimum Pearson χ^2			Minimum logit χ^2		
	Lower bound	Before R-B	After R-B	Lower bound	Before R-B	After R-B
5	137	139	137	131	137	131
6	.142	144	142	.135	.142	135
7	156	162	158	.151	159	.151
.8	194	.211	200	.182	.197	.183
.85	225	255	239	206	224	.206
9	271	.305	282	.239	261	.239

$\beta = 0\,84730$ known, $\alpha = 0$ to be estimated, conditions as in Table 1 Comparison for the χ^2 estimates, of the lower bound values with the mean-square errors before and after Rao-Blackwellization of the estimates. In this table the statistics for the minimum logit χ^2 estimate do not exclude the two samples omitted in Table 1

fore a point which I [5] discussed several years ago in connection with the χ^2 test Very frequently, data as they present themselves permit different ways of properly calculating the χ^2, and this presents difficulties for the interpretation of the results when the χ^2 test is done with any particular arrangement. Now we encounter the difficulty in relation to estimation.

Consider the situation of a die with probability of "success" P, thrown a very large number of times N, or equivalently N identical dice each with characteristic P thrown once, or, more generally still, that there have been N throws accomplished by throwing N/r sets with r in each set. Perhaps, as in Weldon's [21] famous case, they have been thrown 12 at a time. If they are fair, the probability P of the appearance of a 5 or 6 is 1/3, and if the "null hypothesis is true," the expected numbers for 0, 1, 2, $\cdots$, 12 successes in the throws are given by the 13 successive terms of $N/12(2/3+1/3)^{12}$. To test this we can

set the corresponding observed frequencies against the expectations and perform the χ^2 test as for 12 D F. (We suppose N large enough so that we can use the asymptotic χ^2 distribution.) Now someone might not like 12 and propose that the observations should be obtained by taking up the $N/10$ sets of 10 successive dice each, for a test with 10 D.F., or generally, the N/r sets for a test with r degrees of freedom against the expectations given by

$$N/r(2/3 + 1/3)^r.$$

Which of these should be used for the test?[12] Or, to consider the estimation problem, suppose we do not wish to test whether $P=1/3$ but instead we do not know P and wish to estimate its value as $\hat{p}$ by minimizing χ^2; the χ^2 of which arrangement should we minimize?

We can note that if we take $r=1$, we have the four-fold table χ^2, which in fact is the one we have used as the minimum χ^2 estimate; this gives $\hat{p}=S/N$ where S is the total number of successes, and this estimate is the same as the maximum likelihood estimate Now it is known [9] that for a regular unbiased estimate of P, the sampling variance of this estimate is the smallest possible, and that should be a sufficient reason for choosing this χ^2 to minimize, as the definition of the minimum χ^2 estimate, for it is self-defeating to stigmatize the minimum χ^2 estimate as inferior, by reference to a χ^2 estimate that is less than optimum.

We can similarly analyze the minimization of χ^2 for the estimate of the parameter λ of a Poisson distribution. If we have observed frequencies $x_1, \cdots, x_i, \cdots, x_n$ considered hypothetically as a sample of a Poisson distribution with parameter λ

$$p(x_i) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}, \tag{16}$$

the expected numbers for x having the values $0, 1, 2, \cdots, i, \cdots, r$ are given by

$$\hat{f}_i = np(x_i). \tag{17}$$

We may evaluate a χ^2 for a test with r degrees of freedom by enumerating the corresponding observed frequencies of $f_1, f_2, \cdots, f_i, \cdots, f_r$,

$$\chi^2_{(r)} = \sum_{i=1}^{i=r} \frac{(f_i - \hat{f}_i)^2}{\hat{f}_i}. \tag{18}$$

[12] There are many more arrangements possible. For instance, we may add the χ^2's of the $N/10$ sets of 10 for a test based on N degrees of freedom.

The χ^2 of (18) may be modified, if we do not know λ, to evaluate a χ^2 as for $(r-1)$ degrees of freedom by substituting $\hat{f}'$ for $\hat{f}$, with $\hat{f}'$ evaluated from (16) using the mean $\bar{x} = \sum x/n$ instead of λ.

$$\chi^2_{(r-1)} = \sum_{i=1}^{i=r} \frac{(f_i - \hat{f}_i')^2}{\hat{f}_i'} . \tag{18'}$$

If, instead of considering the frequencies f of the frequencies x, we consider the x directly as observed frequencies, and since the expectation for each x is λ, we can evaluate a χ^2 as for n degrees of freedom.

$$\chi^2_{(n)} = \sum_{i=1}^{i=n} \frac{(x - \lambda)^2}{\lambda} . \tag{19}$$

The χ^2 of (19) may be similarly modified, if we do not know λ, by substituting $\bar{x}$ for λ, and evaluating a χ^2 as for $(n-1)$ degrees of freedom.

$$\chi^2_{(n-1)} = \sum_{i=1}^{i=n} \frac{(x - \bar{x})^2}{\bar{x}} . \qquad (19')[13]$$

Consider again the $\chi_{(n)}^2$ of (19) If the observed frequencies x_1 and x_2 are added together and set against the expectation 2λ, this taken together with the remaining $(n-2)$ values of x gives a χ^2 as for $(n-1)$ degrees of freedom. More generally, if r values of x are combined to constitute a group for which the expectation is $r\lambda$, this taken together with the remaining $(n-r)$ values of x for each of which the expectation is λ, gives a χ^2 for $(n-r+1)$ degrees of freedom. If we let $r=n$, we have for the expected total $n\lambda$, and for the observed $\sum^n x = n\bar{x}$ and a χ^2 as for 1 degree of freedom

$$\chi^2_{(1)} = n \frac{(\bar{x} - \lambda)^2}{\lambda} . \tag{20}$$

If the χ^2 of (20) is minimized for the estimate of λ, we obtain

$$\hat{\lambda} = \bar{x}. \tag{21}$$

The estimate (21) is identical with the maximum likelihood estimate of the Poisson parameter, and is known to be "best."

We see then that in the cases considered there are many statistics of the observations distributed as χ^2 (in "large samples"), which yield minimum χ^2 estimates, but only one which yields the "best" estimate.

[13] The χ^2 of (19') is sometimes referred to as the "index of dispersion" of the Poisson distribution.

I suggest that unless otherwise stated we should mean by the "minimum χ^2 estimate" the "best" minimum χ^2 estimate, that is, one obtained by minimizing that χ^2 which yields the "best" estimate. By "best" estimate I should mean the one with smallest mean square error, but some other definition can of course be taken.

In the case of estimating the binomial or Poisson parameter, we recognized which χ^2 is "best," because we knew independently which estimate is best. Can we state some rule by which the "best" χ^2 would be recognized, without this knowledge? Perhaps, but as far as I know this question has not been investigated mathematically, and no rule can be presently given. It is interesting, however, to speculate! Suppose we consider the use of the χ^2's which we can calculate from the observations, for tests of significance instead of for estimation—which χ^2 should then be used? Here I think we must invoke the Neyman-Pearson theory of tests of significance and say, "We should choose the χ^2 which is the most powerful test for a specified alternative—and it would depend on what that alternative is."

If we recall again Weldon's series of dice thrown in sets of 12, we may suppose a situation in which we were quite certain that the dice were all "fair," that is, if they were thrown at random in a long series of throws, the limiting value of the relative frequency of 5 or 6 would be 1/3, but where we were not sure that they were in fact thrown fairly, that is, we thought they might have been manipulated to give favorable combinations. It seems then, on intuition, that it would be well to consider the χ^2 calculated as for $r=12$, because since the dice were thrown in groups of 12, the manipulations would have been performed in this grouping, and the discrepancy from the expected χ^2 would be likely to be apparent for this arrangement, if the r was taken as, say, $r=6$, this discrepancy might be "averaged out"

But suppose that, on the contrary, we knew that each had the same probability P of showing 5 or 6, and that the throws were at "random," but we were uncertain as to whether $P=1/3$. Then I think it can be shown that the most powerful χ^2 test is with the use of the χ^2 with $r=1$. Supposing that this is correct, then we have an identity between the χ^2 to minimize for the estimate of P and the χ^2 to use for a test in which the only admissible alternative is $P\neq 1/3$ This is aesthetically satisfying in that the χ^2 which is best for testing for P is the same as that used for estimating P It suggests itself that this may be the general rule, that is, that the "best" χ^2 for estimate of a parameter θ of which the probability of the observations is a function, is the best χ^2 for testing against an alternative to the tested hypothesis, different only in

respect of the value of the parameter θ. However, this is here quite speculative, and requires for proper elucidation competent mathematical investigation.

APPENDIX NOTE 2. METHODS OF SAMPLING AND CALCULATION

For the main experiment, dealing with the logistic function $\alpha=0$, $\beta=0.84730$, at various arrangements of equally spaced doses, statistics were calculated as for the total population of samples, except in the case of the minimum Pearson χ^2 estimate of the two parameters simultaneously, for which a stratified random sample of 1,000 was used with each dosage arrangement For each of the rest of the experiments, a stratified random sample of 100 was used

The method of constructing the stratified random samples can be described in terms of the experiment with 3 equally spaced doses corresponding respectively to true P equal to 0.1, 0.5, 0.9 We wish to have 100 samples, each representing an exposure of 10 animals at each of the 3 doses The 100 observed values of p for the first dose represent samples from a binomial distribution with true $P=0.1$ With 10 animals exposed there are 11 possible values of p The expected number of samples for each of these, from $p=0$ to $p=1$, for a total number of 100 samples was calculated from the binomial distribution $100(0.9+0.1)^{10}$ For each p a number of cards were counted out equal to the nearest integer with this expected number, and on each card was punched the value of p For the fractional units of expected numbers, cards were assigned to p's at random, to make up the total of 100 When this part of the procedure was finished, we had a sample of 100 very closely representative of the binomial $(0.1+0.9)^{10}$. These were now randomized by punching a random number of 3 or more digits on each card and sorting the 100 cards on this random number.[14] In the same way a stratified random set of p's was obtained for $P=0.5$, and for $P=0.9$. The first p of each of the 3 randomized sets constituted the first of the 100 samples to be used for the experiment, the second p of each set the second sample, and so forth The resulting set of 100 samples is designed to give a better estimate of the desired statistics than would be obtained from a random sample not stratified according to the expected p's

Several comparisons were made of the results obtained with samples constructed in this way and the correct statistics as calculated from the total population, and it was found that a stratified random sample

[14] Ordering on a random number randomizes the original ordered series of cards

as large as 100 gave very good estimates. As examples, there are shown in Tables 8 and 9 comparisons of statistics as obtained from a stratified random sample of 100 and as obtained from calculation for the entire sampling population where these were available, for estimation of α alone and for estimation of both α and β, with central $P=0.5$ and with central $P=0$ 7. It is seen that very satisfactory estimates are obtained with a stratified random sample of as many as 100.

TABLE 8

ESTIMATE OF α, β KNOWN· COMPARISON STRATIFIED SAMPLE AND TOTAL POPULATION

Estimator E	Variance		Ratio max lik./E	
	Total pop.	100 samples	Total pop.	100 samples
Central $P=0$ 5				
Max. lik.	.158	.156	1.00	1.00
Min. χ^2	.139	136	1 14	1.15
Min. logit χ^2	.137	130	1 15	1 20
Central $P=0$ 7				
Max lik.	186	.192	1 00	1 00
Min. χ^2	.161	.161	1 16	1 19
Min logit χ^2	156	155	1 19	1 24

Comparison of variances as for total population and as obtained from 100 stratified random samples $\beta=0.84730$ known, $\alpha=0$ to be estimated, 3 equally spaced doses, 10 at each dose.

For the calculation of the statistics of the total population, a minimum of 10 significant figures was used With the estimates themselves written to 5 decimal places correct within ± 5 in the last figure, the final statistics were written to 3 decimal places and are believed to be generally correct to within ± 1 in the last figure. The estimates obtained with the stratified random samples appear to be generally reliable within ± 5 of the second significant figure.

Punch card machines, including sorter, tabulator, summary punch and collator, but not automatic multiplying punch, were available and were used, as these were found convenient. For the calculation of the estimates, electric Monroe calculators were used. For obtaining the p corresponding to given logit, the exponential tables of reference [25]

were used, and for obtaining the logit corresponding to given p the tables of the natural logarithms of reference [26] were used. In the case of the maximum likelihood and minimum Pearson χ^2 estimates, iterations were continued till the difference between two successive iterations was less than 0 00005, and the estimates, being written to 5 decimal places, were therefore correct within ± 5 of the last figure. Each estimate after calculation was checked by insertion into the con-

TABLE 9

ESTIMATE OF α AND β: COMPARISON STRATIFIED SAMPLE AND TOTAL POPULATION

Estimator E	Estimate of α				Estimate of β			
	Variance		Ratio max. lik./E		Variance		Ratio max lik /E	
	Total pop.	100 samples	Total pop.	100 samples	Total pop.	100 samples	Total pop.	100 samples
Central $P=0$ 5								
Max. lik.	.187	183	1 00	1 00	313	291	1 00	1.00
Min logit χ^2	154	.150	1 21	1 22	.268	254	1.17	1.15
Central $P=0$ 7								
Max lik.	430	.406	1 00	1 00	393	.350	1.00	1.00
Min. logit χ^2	393	.375	1 09	1 08	.274	257	1 43	1 36

Comparison of variances as for total population and as obtained from 100 stratified random samples. $\alpha=0$, $\beta=0$ 84730, both to be estimated, 3 equally spaced doses, 10 at each dose.

ditional equations of estimation. For calculation of the statistics a minimum of 10 significant figures was retained. The estimates of the two main experiments (α to be estimated, and α together with β to be estimated) for each of the 1,331 possible samples and their squares were punched on cards, as well as the probability of the sample, so that sums and sums of squares were obtainable by machine tabulation. A certain amount of over-all checking was done by use of the automatic multiplying punch through contract with an International Business Machines Corporation service bureau All work was checked, in some cases more than once, by direct and indirect checks. Still there is no

guarantee that the results are entirely free from arithmetic error, though I am fairly certain there is no serious mistake in the final results arising from this source.

The estimates themselves were obtained as follows: For the minimum Pearson χ^2 estimate of the two parameters simultaneously, a routine iterative procedure in terms of logits was employed, using weights and working values as given in reference [2], except that here, as for all other estimates requiring iterative procedures, what was solved for in each iteration was ∂a and ∂b, the corrections to provisional values, rather than a and b. For the estimate of the one parameter α with β known, the same method was used, with of course only the one equation instead of two to be solved. When the main work was completed I learned from Dr. William Taylor that he had developed an explicit solution for this case! This is given by

$$a = \frac{1}{2} \ln \frac{\sum p_i^2 e^{-\beta x_i}}{\sum q_i^2 e^{\beta x_i}} \cdot$$

Had I known of Taylor's solution at the start of the experiment, the statistics of the minimum χ^2 estimate of α with β known could have been obtained with less than half the time and effort required by the iterative procedure. Taylor's solution was utilized for a certain amount of checking.

For the maximum likelihood estimates the following general scheme was used, which is appreciably easier than that usually advanced The equations of estimate are, for the present situation with equal n at all doses,

$$\sum (p - \hat{p}) = 0, \tag{22}$$

$$\sum x(p - \hat{p}) = 0. \tag{23}$$

If we have $\hat{p}_0$, a provisional value of $\hat{p}$ corresponding to provisional values a_0 and b_0, and if $\hat{l}_0$ is the corresponding provisional logit, that is, if

$$\hat{l}_0 = a_0 + b_0 x,$$

we can write for $\hat{p} - \hat{p}_0$, approximately

$$\hat{p} - \hat{p}_0 \cong (\hat{l} - \hat{l}_0)\hat{p}_0\hat{q}_0, \tag{24}$$

and from (22) (23) the approximate equations of estimate are

$$\sum[(p - \hat{p}_0) - (\hat{l} - \hat{l}_0)\hat{p}_0\hat{q}_0] = 0, \tag{25}$$

$$\sum x[(p - \hat{p}_0) - (\hat{l} - \hat{l}_0)\hat{p}_0\hat{q}_0] = 0. \tag{26}$$

Since $(\hat{l}-\hat{l}_0)=\partial a+\partial bx$, where ∂a and ∂b are the corrections to the provisional estimates a_0 and b_0, the corrections are given by:

$$\partial b = \frac{\sum px - \sum \hat{p}_0 x - \dfrac{\sum \hat{p}_0\hat{q}_0 x(\sum p - \sum \hat{p}_0)}{\sum \hat{p}_0\hat{q}_0}}{\sum \hat{p}_0\hat{q}_0 x^2 - \dfrac{(\sum \hat{p}_0\hat{q}_0 x)^2}{\sum \hat{p}_0\hat{q}_0}}, \tag{27}$$

$$\partial a = \frac{\sum p - \sum \hat{p}_0 - b\sum \hat{p}_0\hat{q}_0 x}{\sum \hat{p}_0\hat{q}_0}. \tag{28}$$

Again it was not till most of the relevant work was done, that the explicit approximate solution of the maximum likelihood estimate for 3 equally spaced doses, equal n at each dose, advanced by Wilson and Worcester [30], was examined Although these authors put down their published estimates to only 3 significant figures, it was found by trial that their solution usually provides estimates correct to 4 or 5 significant figures. Had this been realized early, a great deal of labor could have been saved in the calculation of the maximum likelihood estimate for the 3 dose experiments.

What has been described in previous paragraphs is the procedure used for obtaining the maximum likelihood estimates in all cases except the experiment with 11 doses, β known, α to be estimated For large number of doses, equal n at each dose, Taylor has developed a method of obtaining a very good approximation of the maximum likelihood estimate explicitly without iterative procedures For estimation of α with β known and with equal numbers at all doses, the maximum likelihood estimate is obtained by equating $\sum \hat{p}$ to $\sum p$ For a large number S of equally spaced doses x, coded successively $0, 1, 2, \cdots, (S-1)$, $\sum \hat{p}$ is given with close approximation by the definite integral

$$\sum \hat{p} \cong \int_{-1/2}^{S-1/2} \frac{1}{1+e^{-(a+bx)}}\, dx, \tag{29}$$

and setting this equal to $\sum p$ gives

$$a = \ln \frac{e^{-\beta S/2} - e^{\beta S/2+\beta(\Sigma p - S)}}{e^{\beta(\Sigma p - S)} - 1}. \tag{30}$$

For the 11 dose experiment the method of Taylor gave the maximum likelihood estimate of α correct to 4 significant figures in almost half the samples and correct to at least 3 significant figures in practically all samples.

It is pertinent to note that the methods outlined in previous paragraphs for calculating the minimum χ^2 and maximum likelihood estimates of the logistic function are considerably easier than are available for the same estimates of the probit equation. There is no explicit solution of the minimum Pearson χ^2 estimate of the mean of the probit equation, with the standard deviation known, corresponding to Taylor's solution for the logistic function, when α is to be estimated with β known nor can Taylor's explicit approximate solution of the maximum likelihood estimate of the logistic function be applied to the probit equation, nor can Wilson and Worcester's Even the iterative procedure used for the maximum likelihood estimate of the logistic function in the general case as outlined in previous paragraphs is appreciably easier than any presently available for the probit equation.

The advantage for calculation with the logistic function can be traced to the fact that with the iterative procedure for obtaining the maximum likelihood estimate, there is required implicitly the calculation of the quantity $\sum(\hat{z}_0/\hat{p}_0\hat{q}_0)(p-\hat{p}_0)$, where $\hat{p}_0$, $\hat{q}_0$, $\hat{z}_0$ represent provisional values which are different from iteration to iteration With the probit equation this involves for each iteration and for each observation, computation of the quantity $(\hat{z}_0/\hat{p}_0\hat{q}_0)p$, but since with the logit equation $\hat{z}_0=\hat{p}_0\hat{q}_0$, the coefficient $\hat{z}_0/\hat{p}_0\hat{q}_0$ is unity and one needs to calculate only $\sum p$, and this only once The same point applies to the quantity $(\hat{z}_0/\hat{p}_0\hat{q}_0)xp$ which is necessary in the estimation of β, replaced in the case of the logistic function by the single computation of $\sum xp$ A secondary advantage arises from the fact that, since the Taylor's series approximation of $(p-\hat{p})$ in the terms of the linear transform is better in terms of the logit transform than in terms of the probit transform [3], the iterations with the logistic function converge somewhat more rapidly.

APPENDIX NOTE 3. THE CASE OF ZERO SURVIVORS

The equations which give the minimum logit χ^2 estimate are

$$\sum n_i p_i q_i l_i = \sum n_i p_i q_i \hat{l}_i, \tag{31}$$

$$\sum n_i x_i p_i q_i l_i = \sum n_i x_i p_i q_i \hat{l}_i, \tag{32}$$

where n_i, $p_i=1-q_i$, and l_i are the number exposed, the observed relative frequency of response, and its logit $\ln p_i/q_i$ respectively, at dose x_i, and $\hat{l}_i$ is the estimated logit given by $\hat{l}_i=a+bx_i$.

For an observed p_i which is either zero or 100 per cent, the value of l_i is infinite, the value of p_iq_i is zero, and the value of $p_iq_il_i$ approaches

zero as p_i approaches either extreme value. If with p observed as zero or 100 per cent, we give $p_iq_il_i$ its limiting value of zero, then in the solution of (31) (32) these observations are effectively eliminated; whether an observation had been made and it turned out to be zero or whether it turned out to be 100 per cent, or whether no observation had been made at all, the estimate derived is the same. On the basis of "common sense," as well as on the basis of mathematical statistical theory, an observation of zero or 100 per cent made at some particular x_i furnishes "information" as respects the estimate, and omitting it wastes such information, if indeed it does not warp the estimate. It may be argued that it is not the experimenter who omits the observation, but the mathematics that does it, and that this endows the omission with statistical validity. Such a view perverts the purpose of statistical estimation, which is not to fulfill a preconceived general principle, but to produce an estimate If there are specific situations in which an estimator behaves badly, it is not only permissible, it seems to me that it is required by sound statistical theory that the estimator be modified or abandoned in that situation, and a supplementary rule of estimation be applied which circumvents the difficulty and yields an estimate which in actual application will be acceptable One such rule for the situation here considered is to use for an observation of zero a replacement working value of $1/2n$, and for an observation of unity a working value $(1-1/2n)$. The reasonableness of such a rule may be set forth as follows:

Consider an experiment in which the samples are drawn at 3 equally spaced doses x_1, x_2, x_3, corresponding to which the observations are $p_1=0.3$, $p_2=0\ 5$, $p_3=0\ 7$, since the logits of these fall on a straight line, we expect that the estimates of the parameters will correspond to this line, and this should be the case with any consistent estimator Consider now another sample with $p_2=0\ 5$ as for the first sample but with p_1 and p_3 decreased and increased respectively by an amount Δp—for instance with $\Delta p=1/10$, $p_1=0\ 2$, $p_3=0\ 8$; from this sample we shall obtain the same estimate of α as with the first sample, but b will be larger. If we change the values of p_1 and p_3 again in the same direction as before, b will increase again, and in general b will increase monotonically with Δp The value of p cannot change continuously, but for any specified n only in steps of $1/n$; the smallest possible value for p greater than zero is $1/n$ and the largest possible value less than unity is $(1-1/n)$ With these values, the estimate of β will be $b=-\text{logit}\ 1/n=\text{logit}\ (1-1/n)$. At the next possible values of p_1 and p_3 with addition of Δp, the estimate of β is $b=\infty$. This value of β is meaningless practically and in fact is in contradiction of the fundamental assump-

tion that the true P's follow a logistic function, for such a value of β implies a true P of zero and unity at finite doses, whereas the logistic function is asymptotic to these values of P.

The monotonic increase of b with Δp, taken together with the exclusion of an estimate of β corresponding to $p_1=0$, $p_3=1.0$, implies an estimate of β that corresponds to an observation p_1, $0<p_1<1/n$, and of p_3, $1>p_3>(1-1/n)$. It seems reasonable to set p_1 halfway between 0 and $1/n$, that is, at $1/2n$, and p_3 halfway between $(1-1/n)$ and 1, that is at $(1-1/2n)$ Any other working values within these limits would meet the requirements of the monotonicity of increase of the estimate of β with Δp and the exclusion of an infinite estimate for β, that is we could take as the working value to replace an observed p of zero, a value $p'=1/kn$ where $k>1$ is some other value than 2, but $k=2$ recommends itself on the basis of simplicity if nothing else

Having adopted the rule of $1/2n$ provisionally, I performed a considerable number of experiments simulating situations of practical bio-assay, to compare the behavior of the estimates obtained in this way with those gotten when other procedures were used Among other procedures tried was one analogous with that used in "probit analysis," where a transform line is fitted to the points, excluding the observations of zero and 100 per cent and using as substitute observations, ones predicted by the fitted line For the minimum logit χ^2 estimate, on the basis of what evidence I have been able to accumulate, I am quite satisfied with the simple $2n$ rule. Using it, all samples yield finite estimates, and judged on the basis of the mean square errors, these estimates were better than those obtained using any other rule which was tried

The arithmetic difficulty of what to do with observations of zero or 100 per cent, when one is using the linear transform of a function and the value of the transform corresponding to these observations is infinite, arose early in the use of the normal deviate or probit for fitting the integrated normal curve The problem was brought to R. A. Fisher, who gave the answer that is incorporated in the now famous Appendix to a paper by C Bliss [8], the title of which I have borrowed for the heading of this section, because it is essentially the same problem as the one with which we are dealing here The answer of Fisher was to use for the transform corresponding to an observed p of 100 per cent the working value $\hat{l}+\hat{q}/\hat{z}$ and for an observed p of zero the working value $\hat{l}-(\hat{p}/\hat{z})$, where $\hat{p}=1-\hat{q}$ is the estimate of the true response at the dose, $\hat{l}$ is the linear transform value corresponding to $\hat{p}$, and $\hat{z}$ can be defined as $\partial\hat{p}/\partial\hat{l}$. For the integrated normal curve, z is the

ordinate of the normal distribution at x; for the logistic function, it is equal to pq. This treatment of zero and 100 per cent response, first given by Fisher for the integrated normal curve, was applied also to the logistic function [17] and judging from a recent declaration [11], is considered by him to be mandatory for these and all similar situations

Not long after the publication of Fisher's [8] note, it was made clear [7] that the substitution of the working value defined for the observations of zero and 100 per cent was the application to these observations of a necessary modification of any observation p, in maximum likelihood estimation using the linear transform, the general formula for the working value being $(p-\hat{p})/\hat{z}+\hat{l}$ This device certainly meets the difficulty in many cases, and a finite estimate is obtained, in spite of the fact that the value of the transform for the observation is infinite It is apparently widely believed that it meets the difficulty in all cases, but I found early in this investigation that it does not

There are samples which do not yield a finite estimate by maximum likelihood for one or both of the parameters to be estimated even using the working value of Fisher. In general, these are samples which, in an experiment with N doses, have N or $(N-1)$ observations of zero or 100 per cent [15] Thompson [29] called attention to the fact that for some such samples the maximum likelihood method does not converge to a finite estimate, and emphasized the necessity to make ad hoc provision for such cases For instance in the estimate of α given β as known, for the 3 dose experiment, samples with all three observations zero per cent or all three 100 per cent response, yield no finite estimate by maximum likelihood. In the estimate of both parameters simultaneously, with the first observation zero per cent, the last observation 100 per cent, and the observation at the second dose 50 per cent, maximum likelihood yields no finite estimate for β, and with any other observation at the central dose, it yields no finite estimate for either parameter In some situations these insoluble samples are extremely rare, in others they are not rare, while in some situations they are extremely frequent For instance, in an experiment with 3 equally spaced doses, the corresponding true P's of which are 0 1, 0 5, 0 9, the insoluble samples constitute more than 12 per cent of the total sampling population; in an experiment with 4 equally spaced doses the lowest of which corresponds to true $P=0\ 01$ and the highest of which corresponds to $P=0\ 99$, more than 20 per cent of the samples are insoluble by maximum likelihood If in the 3 dose experiment the P of the lowest dose is P

[15] But not necessarily any such sample The same is true of the minimum Pearson χ^2 estimate

$=0.01$ and for the highest dose it is $P=.99$, more than 80 per cent of the samples are insoluble by maximum likelihood.[16]

I do not see how these estimates can be considered acceptable. The situations referred to are not "pathological." On the contrary, they are quite reasonable; some of them correspond to very good experimental designs and yield excellent estimates by other available methods of estimation An estimator that yields an infinite estimate in from 10 to 80 per cent of samples for some well-designed experiments can hardly be considered "proper" and it seems to me that one is obliged either to modify or to abandon it

I have not had the temerity to modify the maximum likelihood or minimum Pearson χ^2 estimate as I have done with the minimum logit χ^2 estimate. In the text, the comparisons of the mean square error of the minimum logit χ^2 estimates with the other estimates are made, omitting, in the calculation of the statistics of all estimates, samples which had an infinite estimate by maximum likelihood For the main experiment with dosages centrally placed, these samples are not very frequent, but as the disposition of the dosages is moved to positions asymmetrical with respect to central $P=0.5$, insoluble samples become more frequent and the comparisons referred to in the text do not include situations in which samples insoluble by maximum likelihood constitute as much as five per cent of the total

The question has been raised as to whether the comparisons showing results advantageous to the minimum logit χ^2 estimate are not the consequence solely of the application of the $2n$ rule to the minimum logit χ^2 estimates, and not to the maximum likelihood estimates As respects this question, it should be noted first that if the $2n$ rule did in fact account for the better results where these were obtained, and this was not peculiar to some special situation, it would be no adverse reflection on the results, but rather a support of the soundness of the $2n$ rule. However, it is not a fact that the application of the $2n$ rule to the minimum logit χ^2 estimate and not the maximum likelihood estimate accounts for all the differences found between the estimates.

[16] The behavior of the maximum likelihood estimator in the present situation strikes one as logically obstreperous If we think of the principle of maximum likelihood in its general intent, it should estimate for a parameter θ, a value which, if it were the true value, would yield the sample in hand most frequently One would suppose, then, that if a sample S is very frequent with some particular θ, such a sample would characteristically yield θ by maximum likelihood, or estimates near θ. Now, with a 3 dose symmetrical experiment, 10 at each dose, with $\alpha=0$, $\beta=4.595$, samples in the class 0, p, 1.0, where p is any value $0<p<1.0$, occur with probability about 80 per cent, yet no such sample yields by maximum likelihood any value near $\beta=4.595$, for all yield $\beta=\infty$, and only one sample in this class yields a finite estimate for α. On the other hand, samples in the class 0, 1.0, p occur only with probability of 0.09 per cent, yet with all these samples except 0, 1.0, 1.0, maximum likelihood indicates finite estimates for α and β, not excessively far from the true values.

Even if the $2n$ rule is applied to the observations of zero or 100 per cent for the maximum likelihood estimate as well as for the minimum logit χ^2 estimate, the comparisons favorable to the minimum logit χ^2 estimates remain. In Table 10 is shown a comparison of statistics of the maximum likelihood and minimum logit χ^2 estimates, with the $2n$ rule used in the case of observations of zero or 100 per cent response for

TABLE 10

APPLICATION OF $2n$ RULE TO MAXIMUM LIKELIHOOD ESTIMATE

	Mean-square-error					
	Central $P = 5$			Central $P = .7$		
	Min logit χ^2	Max. lik unmodified	Max lik. $2n$ rule	Min. logit χ^2	Max lik. unmodified	Max. lik. $2n$ rule
Estimate of α, β known						
a	.130	156	.151	.158	193	.171
Estimate of α and β						
a	.150	183	.171	.375	.406	.394
b	256	.299	287	.257	.361	.275

Mean square errors with application of the $2n$ rule for maximum likelihood estimates Based on 100 stratified random samples The maximum likelihood estimates are improved, but the relative position of the estimates is unchanged

the maximum likelihood estimate as well as for the minimum logit χ^2. It is seen that while the results are better for the maximum likelihood estimate than when Fisher's rule is applied (Table 2), they still are inferior to those obtained with the minimum logit χ^2 estimate

APPENDIX NOTE 4. SUFFICIENCY OF THE ESTIMATES

A function of the observations T is a sufficient statistic for a parameter θ if, with respect to any other statistic T', the conditional probability of T' given T is independent of θ. If ϕ is the probability of a sample set of observations, and T is a sufficient statistic, then for all samples with the same value of T, the value of $\partial \ln \phi/\partial\theta$ is the same, and this being true, if it is also true that for all samples having the same value

of $\partial \ln \phi / \partial \theta$, the value of T is the same, T is a minimal sufficient statistic. In general, a sufficient statistic is minimal sufficient, if it is a function of any other sufficient statistic [19a].

For the logistic function

$$\frac{\partial \ln \phi}{\partial \alpha} = \sum n_i(p_i - P_i), \tag{33}$$

where n_i is the number "exposed" at x_i, p_i is the observed relative frequency at x_i, and P_i is the true value of the probability at x_i given by the logistic function Since n_i and P_i are constants from sample to sample, the value of (33) will be the same in samples, if the value of $\sum n_i p_i$ is the same in the samples It follows that $\sum n_i p_i$ is a sufficient statistic for α in the present situation, and it is a minimal sufficient statistic

Similarly,

$$\frac{\partial \ln \phi}{\partial \beta} = \sum n_i x_i (p_i - P_i) \tag{34}$$

and $\sum n_i x_i p_i$ is a minimal sufficient statistic for β.

A more direct demonstration of the sufficiency of the statistics $\sum n_i p_i$, $\sum n_i x_i p_i$ can be given as follows·

If $P_i = 1 - Q_i$, n_i, $r_i = p_i n_i$, $s_i = n_i - r_i = q_i n_i$, are respectively the probability of response, the number exposed, the number responding, and the number not responding, at x_i, the probability of a sample is

$$\begin{aligned} \phi &= \prod C^{n_i}{}_{r_i} P^{r_i}{}_i Q^{s_i}{}_i \\ &= \prod C^{n_i}{}_{r_i} \frac{\left[e^{-(\alpha+\beta x_i)}\right]^{s_i}}{\left[1 + e^{-(\alpha+\beta x_i)}\right]^{n_i}} \\ &= \prod C^{n_i}{}_{r_i} \prod \left[1 + e^{-(\alpha+\beta x_i)}\right]^{-n_i} e^{-\alpha \Sigma s_i} e^{-\beta \Sigma s_i x_i} \\ &= \prod C^{n_i}{}_{r_i} \prod \left[1 + e^{-(\alpha+\beta x_i)}\right]^{-n_i} e^{-\alpha \Sigma n_i - \beta \Sigma n_i x_i} e^{\alpha \Sigma r_i} e^{\beta \Sigma r_i x_i}. \end{aligned} \tag{35}$$

It can be seen directly from (35) that if α is changed by an amount $\Delta\alpha$, all samples having the same value for the statistic $\sum r_i$ will be changed in probability by the same factor and therefore that the frequency distribution of these samples will remain unchanged Hence $\sum r_i = \sum n_i p_i$ is a sufficient statistic for α Similarly, if β is changed by an amount $\Delta\beta$, all samples having the same value for the statistic $\sum r_i x_i$ will be changed in probability by the same factor, and the fre-

quency distribution of the samples will be unchanged. Hence $\sum r_i x_i = \sum n_i p_i x_i$ is a sufficient statistic for β.

It is evident that any one-to-one function of a sufficient statistic is sufficient, and also that if an estimator yields a different value of the estimate for each possible sample, it is sufficient.[17]

Considering joint estimation of α and β, so far as disclosed by this investigation, the minimum logit χ^2 estimate is sufficient, for each possible sample yields a different pair of estimates For the estimate of α given β, at least for all experiments with equally spaced dosages x, 10 exposed at each dose, the minimum logit χ^2 estimate is sufficient, for whenever two or more samples yield the same value for the estimate of α, these samples have the same value for the corresponding minimal sufficient statistic $\sum n_i p_i$

The maximum likelihood estimate is infinite for some samples (See Appendix note 3), and for these the values of the corresponding minimal sufficient statistic $\sum n_i p_i$ or $\sum n_i x_i p_i$ are not always the same;[18] therefore the maximum likelihood estimate is not fully sufficient. However, excluding samples with infinite estimate, there is a one-to-one relation between the maximum likelihood estimate and the corresponding minimal sufficient statistic, and therefore so far as samples with finite estimates are concerned, the maximum likelihood estimate is minimal sufficient.

The minimum Pearson χ^2 estimator yields infinite estimates for the same samples for which the maximum likelihood estimate is infinite and therefore it also fails of full sufficiency For samples yielding finite estimates, the minimum Pearson χ^2 is here sufficient, since all samples yielding the same value of the estimate have the same value of the corresponding minimal sufficient statistic

Some of the findings of this investigation appear to conflict with generalities which are widely accepted as having been established mathematically, for example:

1 A sufficient estimator when it exists is unique, that is, it is either the maximum likelihood estimate or a function of the maximum likelihood estimate [19]

---

[17] As Fisher [15] puts it, " . the intrinsic accuracy of T can never be greater than when every possible sample yields a different value of T in such a case, the actual sample can be reconstructed without ambiguity from the value of T, and so the value of T must contain the whole of the information supplied by the sample "

[18] For instance, all samples 0, 0, p with $0 \leq p \leq 1$ yield an infinite maximum likelihood estimate for α, and all such samples except the sample with $p = 0$ yield an infinite estimate for β, but the value of $\Sigma n_i p_i$ and of $\Sigma n_i x_i p_i$ is different for each such sample

2. Among asymptotically efficient estimators, the maximum likelihood estimator extracts the most information from the data, and where there is a sufficient statistic, it extracts 100 per cent [14].

The first of these generalities is refuted by the finding that the minimum logit χ^2 estimate is sufficient, while it is not a function of the maximum likelihood estimate. Assemble all 1,331 possible samples in groups, such that the samples in each group have the same value for the minimal sufficient statistic of α, $\sum n_i p_i$; we will call such a group of samples a "sufficiency group." The maximum likelihood estimate of α with β known is the same for each sample in a sufficiency group, but the minimum logit χ^2 estimate is not, and in fact is usually, but not always, different for each sample. Fixing the maximum likelihood estimate does not determine the minimum logit χ^2 estimate, and the minimum logit χ^2 estimate is therefore not a function of the maximum likelihood estimate. However, no two samples in different sufficiency groups have the same minimum logit χ^2 estimate, which, since $\sum n_i p_i$ is minimal sufficient, is the necessary and sufficient condition for the sufficiency of the minimum logit χ^2 estimate.

As regards the second generality, the loss of extractable information associated with an estimator can be calculated as $\sum \Phi\sigma^2$, where σ^2 is the variance of $\partial \ln \phi/\partial\theta$ for a set of samples yielding the same value of the estimate, ϕ being the probability of a sample, θ the parameter estimated, and Φ the probability of the set of samples For a three-dose experiment with true P's respectively at 0 01, 0 5, 0.99, both parameters to be estimated, 62 per cent of the population of samples have a maximum likelihood estimate for α which is infinite Moreover these samples do not all have the same value for $\partial \ln \phi/\partial\alpha = \sum n_i(p_i - P_i)$, and the calculated loss of information by the maximum likelihood estimator is 76 0 per cent. For β, 82 per cent of the population of samples yield an estimate of infinity, but most of the samples have the same value for $\partial \ln \phi/\partial\beta = \sum n_i x_i(p_i - P_i)$, and the calculated loss of information by the maximum likelihood estimator for β is only 0 097 per cent. For the minimum logit χ^2 estimate, there are no samples with estimate of infinity, and for all samples yielding the same estimate, the value of $\partial \ln \phi/\partial\theta$ is the same, so there is no loss of information with the minimum logit χ^2 estimate in this experiment.

These calculations and the statements based upon them are made from my best understanding of how to compute the loss of information, but I do not feel entirely secure in the correctness of my procedures. Perhaps Sir Ronald Fisher would object to grouping all samples having infinity as estimate, as having the same value for the estimate; perhaps

different ∞'s must be considered as different values But this poses starkly the conundrum of how these estimates are to be treated, not only as regards practical situations, but also as a mathematical question.

REFERENCES

[1] Berkson, Joseph, "A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function," *Journal of the American Statistical Association*, 48 (1953), 565–99.

[2] Berkson, Joseph, "Minimum χ^2 and maximum likelihood solution in terms of a linear transform, with particular reference to bio-assay," *Journal of the American Statistical Association*, 44 (1949), 273–78.

[3] Berkson, Joseph, "Approximation of chi-square by 'probits' and by 'logits'," *Journal of the American Statistical Association*, 41 (1946), 70–4

[4] Berkson, Joseph, "Application of the logistic function to bio-assay," *Journal of the American Statistical Association*, 39 (1944), 357–65

[5] Berkson, Joseph, "Some difficulties of interpretation encountered in the application of the chi-square test," *Journal of the American Statistical Association*, 33 (1938), 526–36.

[6] Blackwell, David, "Conditional expectation and unbiased sequential estimation," *Annals of Mathematical Statistics*, 18 (1947), 105–10

[7] Bliss, C I, "The determination of the dosage-mortality curve for small numbers," *Quarterly Journal of Pharmacy and Pharmacology*, 11 (1938), 192–216.

[8] Bliss, C I, "The calculation of the dosage-mortality curve," Appendix: "The Case of Zero Survivors," by R A Fisher, *Annals of Applied Biology*, 22 (1935), 134–67.

[9] Cramér, Harald, *Mathematical Methods of Statistics*, Princeton, Princeton University Press (1946)

[10] Cramér, Harald, *Ibid* See p. 480.

[11] Fisher, R. A., "The analysis of variance with various binomial transformations," *Biometrics*, 10 (1954), 130–39.

[12] Fisher, R A., "Professor Karl Pearson and the method of moments," *Annals* of *Eugenics*, 7 (1937), 303–18. See p 317.

[13] Fisher, R. A , "The logic of inductive inference," *Journal of the Royal Statistical Society*, 98, Part I (1935), 39–54 See p 41.

[14] Fisher, R A , "Theory of statistical estimation," *Proceedings of the Cambridge Philosophical Society*, 22, Part 5 (1925), 700–25 See pp. 720–22.

[15] Fisher, R A , *Ibid.* See p. 717.

[16] Fisher, R. A , "On the mathematical foundations of theoretical statistics," *Philosophical Transactions of the Royal Society of London*, Series A, 222 (1922), 309–68

[17] Fisher, R. A., and Yates, F , *Statistical Tables for Biological, Agricultural and Medical Research. Second Edition*, London and Edinburgh, Oliver and Boyd (1943). See pp. 8–11.

[18] Katz, Leo, "On the relative efficiencies of BAN estimates," *Annals of Mathematical Statistics*, 21 (1950), 398–405.

[19] Kendall, M. G., *The Advanced Theory of Statistics*, London, Charles Griffin and Co , Ltd. Vol. II (1946). See chapter 17.

[19a] Lehmann, E. L., and Scheffé, Henry, "Completeness, similar regions and unbiased estimation," *Sankyā*, 10 (1950), 305–39.

[20] Neyman, Jerzy, "Contribution to the theory of the χ^2 test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, edited by Jerzy Neyman, Berkeley and Los Angeles, University of California Press (1949), pp. 239–273

[21] Pearson, Karl, "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling," *Philosophical Magazine*, Series 5, 50 (1900), 157–75.

[22] Rao, C R., *Advanced Statistical Methods in Biometric Research*, New York, John Wiley and Sons (1952) See p 132

[23] Rao, C. R., *Ibid* See pp 130–35.

[24] Rao, C. R., "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, 37 (1945), 81–91

[25] *Tables of Exponential Function e^x* National Bureau of Standards, Washington, D C, U S Government Printing Office (1947)

[26] *Tables of Natural Logarithms*, Vol. 3, Federal Works Agency, Works Projects Administration, Arnold N. Lowan, technical director, National Bureau of Standards (1941).

[27] Taylor, W F., "Consistency of estimators under a specialized bio-assay procedure," Abstract, *Journal of the American Statistical Association* In press

[28] Taylor, W F, "Distance functions and regular best asymptotically normal estimates," *Annals of Mathematical Statistics*, 24 (1953), 85–92

[29] Thompson, W R., "Use of moving averages and interpolation to estimate median-effective dose," *Bacteriological Reviews*, 11 (1947), 115–45. See pp. 139–40

[30] Wilson, E B., and Worcester, Jane, "The determination of L D 50 and its sampling error in bio-assay, II," *Proceedings of the National Academy of Sciences*, 29 (1943), 114–20.

SOME STATISTICAL PROBLEMS IN RELATING EXPERIMENTAL DATA TO PREDICTING PERFORMANCE OF A PRODUCTION PROCESS

T. W Anderson
Columbia University

The performance of a production process is characterized by the quantity and quality of the output, it is affected by factors of production, such as conditions of the process and quantities of inputs In experimentation where the factors are controlled, we assume a bivariate linear regression model with quantity and quality of output as dependent variates In operation where quality of output is controlled by adjusting one of the factors, another regression model is used in which quantity of output and the one factor are dependent variates. The second model is derived from the first It is shown how to use experimental data to estimate the coefficients of the regression of quantity of output in the second model; this regression function is desired for predicting performance in operation. Confidence regions and tests of hypotheses are treated The exposition is in the form of an analysis of a particular problem met by a chemical engineering firm.

1. INTRODUCTION

In this paper we analyze a statistical problem met by a chemical engineering firm (The M W Kellogg Company, Jersey City, N J.) in applying results from experimentation to predicting performance in production The feature of this problem that is of special interest is that in experimentation the two most relevant characteristics of the product, measures of quantity and quality, are permitted to vary but various factors are controlled, while in production one of the product characteristics, quality, is specified and one of the factors is varied to obtain the specified quality The data from the experiments are used to estimate the coefficients of the regression model with two dependent variates which is the model appropriate to experimentation. From this information we derive estimates of the linear function which is appropriate to predicting the quantity produced in commercial operation when quality is controlled by varying a given factor of production. Confidence regions are given, and tests of hypotheses are considered

The particular problem analyzed here arose in work done by the Petroleum and Chemical Research Laboratory of The M. W. Kellogg Company. It occurred in a study of the effect of certain operating

variables on yields and qualities from a refining process involving the solvent extractions of a petroleum fraction for the production of a raw lubricating oil base stock. The paper is written in terms of this specific case, but the procedures used are described in sufficient generality to be applied to other problems which involve the same feature of interchanging a dependent and an independent variable. It is expected that this feature will arise in many other situations; as a matter of fact, in The M. W. Kellogg Company this problem has arisen a number of times The author is indebted to The M. W. Kellogg Company for calling this problem to his attention, for supporting a part of his work, and for furnishing the data analyzed in this paper.

It turns out that the models and statistical methodology used in this problem are similar to those used in some econometric problems involving "structural equations," although the reason that the models are appropriate here are different from the reasons in econometrics.

2. THE ENGINEERING PROBLEM

In the petroleum industry there are various processes in which oil of some kind is fed into a complicated and expensive apparatus; the main product of some processes is oil of a higher quality. In the process with which we are concerned two other liquids besides oil are fed into an extraction apparatus (called an "extractor"); we shall call these "Liquid A" and "Liquid B" Besides the output of desired high quality oil, called "raffinate," there is also drawn from the apparatus an undesired product, called "extract" The flows of liquids A and B relative to the flow of feed oil, of course, affect the quantity and quality of the desired raffinate; the other important characteristic of the process is the temperature. Increasing temperature or flow of liquid A increases the purity of the product, but decreases the yield; increasing the flow of liquid B decreases the purity, but increases the yield The feed oil, and liquids A and B flow into the extractor at constant rates while the process goes on; the flow of liquid A is measured in terms of the volume relative to the volume of oil, the flow of liquid B is measured as the per cent by weight of flow of liquid A. The temperature, measured in degrees Fahrenheit, is actually an average of temperatures in various parts of the extractor There are, of course, a number of other characteristics of the process, but these are relatively unimportant compared to the ones mentioned above, and in the experiments these other characteristics were held as constant as possible The flow of the desired product, namely the raffinate, is measured relative to the flow of the feed. The purity or quality of the raffinate is measured in a way too

complicated to describe here; it is an index with high values desired.

In a given experiment the flows of liquids A and B and the temperature of the process are set, and the flow and quality of raffinate are measured. The extractor is run for several hours; the data used are averages over the period of time for which the process is "stationary" or "in control." For given flows of liquids A and B and for given temperature, the yield and quality vary from experiment to experiment. A part of this variation is due to variation in the characteristics of the feed, though there are other causes of the variation.

When the process is used commercially, it is essential to control the quality of the raffinate This is done by adjusting the temperature of the process so that the raffinate flowing out has a specified purity. The flows of liquids A and B are not varied, but are predetermined Thus in the experiments flows of A and B and temperature are controlled (or "fixed") variables, and yield and quality are uncontrolled, while in production flows of A and B and quality of output are controlled, while yield and temperature are not.

From the experimental data on a given extractor, the engineer wishes to describe the performance of the extractor in production. He wants to be able to predict the yield of raffinate of specified purity that will result from given flows of A and B

3. THE STATISTICAL MODELS

In experimentation the temperature, say x_1, the flow of A, say x_2, and the flow of B, say x_3, are independent variates, and the yield, say y_1, and the purity index, say y_2, are dependent variables We write

$$(1) \quad y_1 = \alpha_{10} + \alpha_{11}(x_1 - \bar{x}_1) + \alpha_{12}(x_2 - \bar{x}_2) + \alpha_{13}(x_3 - \bar{x}_3) + u_1,$$

$$(2) \quad y_2 = \alpha_{20} + \alpha_{21}(x_1 - \bar{x}_1) + \alpha_{22}(x_2 - \bar{x}_2) + \alpha_{23}(x_3 - \bar{x}_3) + u_2,$$

where $\bar{x}_1$, $\bar{x}_2$, and $\bar{x}_3$ are some convenient values of x_1, x_2, and x_3, respectively, u_1 and u_2 are discrepancies and are considered as random or statistical variables The engineers and statisticians working on these problems consider that the linear equations (1) and (2) are adequate; in particular, that u_1 and u_2 can be taken to have a bivariate distribution, approximately normal with zero means and constant variances say σ_1^2 and σ_2^2 and covariance $\sigma_1\sigma_2\rho$ (or correlation ρ). Hence, (1) and (2) are treated as customary regression equations with three independent variates and two dependent variates.

In production x_1 is adjusted to give the desired y_2; that is

$$(3)\qquad \begin{aligned}(x_1 - \bar{x}_1) = \frac{1}{\alpha_{21}} y_2 - \frac{\alpha_{20}}{\alpha_{21}} - \frac{\alpha_{22}}{\alpha_{21}} (x_2 - \bar{x}_2) \\ - \frac{\alpha_{23}}{\alpha_{21}} (x_3 - \bar{x}_3) - \frac{1}{\alpha_{21}} u_2.\end{aligned}$$

We consider u_2 as a discrepancy that holds constant for the entire run. The temperature is varied at the beginning of the run so that the purity is at the desired level, and then the temperature is held constant for the rest of the run. Thus in (3), y_2, x_2, and x_3 are controlled or fixed variables and x_1 is a statistical or dependent variable When we put $x_1 - \bar{x}_1$ as given in (3) into (1) we get

$$(4)\qquad y_1 = \beta_0 + \beta y_2 + \beta_2(x_2 - \bar{x}_2) + \beta_3(x_3 - \bar{x}_3) + v,$$

where

$$(5)\qquad \beta_0 = \alpha_{10} - \frac{\alpha_{11}}{\alpha_{21}} \alpha_{20} = \alpha_{10} - \beta\alpha_{20},$$

$$(6)\qquad \beta = \frac{\alpha_{11}}{\alpha_{21}},$$

$$(7)\qquad \beta_2 = \alpha_{12} - \frac{\alpha_{11}}{\alpha_{21}} \alpha_{22} = \alpha_{12} - \beta\alpha_{22},$$

$$(8)\qquad \beta_3 = \alpha_{13} - \frac{\alpha_{11}}{\alpha_{21}} \alpha_{23} = \alpha_{13} - \beta\alpha_{23},$$

$$(9)\qquad v = u_1 - \frac{\alpha_{11}}{\alpha_{21}} u_2 = u_1 - \beta u_2.$$

In (4) y_2, x_2, and x_3 are fixed, and y_1 is allowed to vary The discrepancy v has mean and variance

$$(10)\qquad \mathcal{E}v = \mathcal{E}(u_1 - \beta u_2) = 0,$$

$$(11)\qquad \mathcal{E}v^2 = \mathcal{E}(u_1 - \beta u_2)^2 = \sigma_1^2 - 2\beta\sigma_1\sigma_2\rho + \beta^2\sigma_2^2.$$

To predict the performance of the extraction process in production one needs to know the coefficients, β_0, β, β_2, and β_3 and the variance In this model, (4) and the variance of v give a complete description of the process.

The statistical procedure we shall use involves estimating the coefficients of (1) and (2) by least squares and estimating σ_1^2, σ_2^2 and ρ by the residuals. The estimates of these parameters are substituted

into (5), (6), (7), (8), and (11) to obtain estimates of β_0, β β_2, β_3 and the variance of v. We shall also find a confidence interval for β and confidence regions for pairs of coefficients such as β and β_3.

The statistical methods applied here have been developed in considerable generality in earlier publications ([1], [2]). However, the methods needed in the present problem will be developed in this paper. Since the methods were originally motivated by considerations of econometric models, so-called "structural equations," it is useful to show that the models given above are identical with those used in econometric models Suppose y_1 and y_2 are "endogenous" variables, that is, economic variables whose formation the model is supposed to describe. Let x_1, x_2, and x_3 be "exogenous" variables; that is, variables whose values are determined by some outside mechanism Then an econometric model would consist of two structural equations of the form $\phi_{i1}y_1+\phi_{i2}y_2=\theta_{i0}+\theta_{i1}(x_1-\bar{x}_1)+\theta_{i2}(x_2-\bar{x}_2)+\theta_{i3}(x_3-\bar{x}_3)+w_i$, $i=1, 2$ $(\phi_{11}\phi_{22}-\phi_{12}\phi_{21}\neq 0)$ Solution of this pair of equations gives two equations (1) and (2), this pair is called the "reduced form." If one of the structural equations, say the first, has coefficients $\phi_{11}=1$ and $\theta_{11}=0$, then it is algebraically identical with (4) In an econometric model this structural equation would have some special economic significance. If the economic system described by the model is changed in a particular way, y_2 might be fixed, and then under this "policy change," (4) would describe the formation of y_1 with y_2 fixed For an elementary exposition of these ideas the reader of this journal is referred to [4].

The statistical methods used in this paper are called in the econometric literature "reduced form" or "limited-information maximum likelihood" methods If the experimental data were obtained in the same way that the extractor would be used in production (that is, with y_2 fixed), then one could use least squares procedures directly on (4). However, since in the experiments y_2 is a dependent variable, the least squares procedure applied to (4) would give biased estimates, and one cannot use the usual confidence intervals and regions appropriate to the usual least squares model. The advantages and disadvantages of these alternative methods have been discussed considerably in the econometric literature; we shall not go into them here.

A natural question to ask is why are the experiments conducted with temperature fixed instead of conducting them with quality controlled One answer is that it seems more "natural" to control temperature since controlling quality is a kind of secondary operation, that is, in the process it is more natural to consider (1) and (2) as the structural equations (as actually reflecting the chemical and physical reaction).

Secondly, the experiments are conducted to obtain more information than just what is necessary to predict yield when quality is controlled by varying temperature; for instance, one might consider controlling quality in production by varying the flow of one of the liquids Moreover, there are other engineering characteristics to be studied.

There is also a relationship between the considerations of this paper and a point raised by Berkson [3] Suppose one considers a pair of variables, measured with "error" but with the "true" parts satisfying a linear relationship; that is, the true parts satisfy $\zeta_1 = c + d\zeta_2$, where ζ_1 and ζ_2 are the true variables and the observed variables are $z_1 = \zeta_1 + v_1$ and $z_2 = \zeta_2 + v_2$, where v_1 and v_2 are errors of observation. Berkson notes (see also [5]) that if z_2 is controlled then $z_1 = \zeta_1 + v_1 = c + d(z_2 - v_2) + v_1 = c + dz_2 - dv_2 + v_1$, and since z_2 is fixed (or, in other terms $v_1 - dv_2$ is statistically independent of z_2) one can consider the above equation as a regression equation. This is analogous to (4) Here is another case where the nature of the model is changed by fixing a variable that was alternatively not fixed.

4. THE EXPERIMENTAL DATA

The data we have at our disposal were obtained from a well-designed experiment or set of experiments, at The M. W Kellogg Company and hence are easy to analyze.

Two levels of each of the controlled variables were used In the table below we write $T=1$ and 2 to mean temperatures of 160° and 180°, $A=1$ and 2 to mean flows of liquid A at 2 0 and 3 5 times the flow of feed (relative volumes) and $B=1$ and 2 to mean flows of liquid B at 2% and 5% of the flow of liquid A (by weight) The first experiment consisted of 16 runs In these 16 runs four different arrangements or patterns of the apparatus were used, indicated as $P=1, 3, 4, 5$ Eight additional runs were made using patterns 3 and 4. We are taking pattern differences to be small enough to be ignored As a matter of fact, an analysis of y_1, for example, for the first 16 runs indicates the pattern differences to be small; the F-value is actually less than one.

Another feature of the data that shows some departure from the assumed model is that the variation in the later runs is less than that in the earlier runs; this is shown by a comparison of the error variances of the last 8 runs. This feature may be due in part, at least, to increasing skill on the part of the engineers in keeping the process "in control." It would have been preferable, of course, to have had the order of runs randomized.

TABLE

RESULTS OF 24 RUNS OF AN EXTRACTOR OF THE M. W. KELLOGG COMPANY

Run. No.	P	T	A	B	Yield	Quality
1	1	2	2	2	54 0	99.4
2	1	2	1	1	49.8	99.0
3	1	1	2	1	55 2	100.4
4	1	1	1	2	69.3	92.6
29	5	2	2	1	50.0	103 3
30	5	2	1	2	69.0	93 6
31	5	1	2	2	63.8	98.2
32	5	1	1	1	64 8	93.4
9	3	2	2	1	47.0	106.0
10	3	2	1	2	55.6	99.4
11	3	1	2	2	61.6	98.8
12	3	1	1	1	64.5	93 5
13	4	2	2	2	52.4	103 5
14A	4	2	1	1	55 7	96.8
15	4	1	2	1	57 8	101.9
16	4	1	1	2	65.9	94.1
21	3	2	2	2	56 9	101.2
22	3	2	1	1	55.7	98.2
23	3	1	2	1	51.3	101.5
24	3	1	1	2	70.6	93.7
25	4	2	2	1	46 2	106 3
14	4	2	1	2	64 1	96 1
27	4	1	2	2	60.7	100 9
28	4	1	1	1	62.8	96 4

5 ESTIMATION OF COEFFICIENTS, VARIANCES, AND CORRELATION COEFFICIENT

Since only two levels of each factor are used, equations (1) and (2) could be written in the form used in analysis of variance by letting $\alpha_{i1}(x_1 - \bar{x}_1)$ be λ_{i2} for $x_1 = 180°$ and be $\lambda_{i1} = -\lambda_{i2}$ for $x_1 = 160°$, letting $\alpha_{i2}(x_2 - \bar{x}_2)$ be μ_{i2} for $x_2 = 3\ 5$ and be $\mu_{i1} = -\mu_{i2}$ for $x_2 = 2\ 0$ and letting $\alpha_{i3}(x_3 - \bar{x}_3)$ be ν_{i2} for $x_3 = 5.0$ and be $\nu_{i1} = -\nu_{i2}$ for $x_3 = 2\ 0$ Then

$$y_i = \alpha_{i0} + \lambda_{ij} + \mu_{ik} + \nu_{il} + u_i, \qquad i = 1, 2, \tag{12}$$

for levels j of x_1, k of x_2, and l of x_3. We shall, however, hold to the original notation because it is more general and hence more suitable for other problems of this sort.

The estimates of the coefficients in (1) and (2) are obtained by the usual methods of univariate regression as

$$a_{ij} = \frac{1}{c_j} \sum_{\gamma=1}^{N} y_{i\gamma}(x_{j\gamma} - \bar{x}_j), \qquad j = 1, 2, 3, \tag{13}$$

where $\gamma(=1, \cdots, N)$ denotes the observation number, $N=24$, $\bar{x}_i = \sum_\gamma x_{i\gamma}/N$, and

$$c_j = \sum_{\gamma=1}^{N} (x_{j\gamma} - \bar{x}_j)^2. \tag{14}$$

Because of the orthogonality properties of the design $\sum(x_{j\gamma}-\bar{x}_j)(x_{k\gamma}-\bar{x}_k)=0$ Finally $a_{i0}=\bar{y}_i=\sum_\gamma y_{i\gamma}/N$. The variance of u_i is estimated from

$$\begin{aligned}(N-4)s_i^2 &= \sum_{\gamma=1}^{N} [y_{i\gamma} - a_{i0} - a_{i1}(x_{1\gamma} - \bar{x}_1) \\ &\qquad - a_{i2}(x_{2\gamma} - \bar{x}_2) - a_{i3}(x_{3\gamma} - \bar{x}_3)]^2 \\ &= \sum_{\gamma=1}^{N} y_{i\gamma}^2 - Na_{i0}^2 - c_1a_{i1}^2 - c_2a_{i2}^2 - c_3a_{i3}^2.\end{aligned} \tag{15}$$

The correlation between u_1 and u_2 is estimated from

$$\begin{aligned}(N-4)rs_1s_2 &= \sum_{\gamma=1}^{N}\left[y_{1\gamma} - a_{10} - \sum_{j=1}^{3} a_{1j}(x_{j\gamma} - \bar{x}_j)\right] \\ &\qquad \cdot \left[y_{2\gamma} - a_{20} - \sum_{j=1}^{3} a_{2j}(x_{j\gamma} - \bar{x}_j)\right] \\ &= \sum_{\gamma=1}^{N} y_{1\gamma}y_{2\gamma} - Na_{10}a_{20} - \sum_{j=1}^{3} c_ja_{1j}a_{2j}\end{aligned} \tag{16}$$

The estimates are

$$\begin{aligned} a_{10} &= 58\ 529, & a_{20} &= 98\ 675, \\ a_{11} &= -\ .3829, & a_{21} &= 1558, \\ a_{12} &= -\ 5.050, & a_{22} &= 4.144, \\ a_{13} &= 2.308, & a_{23} &= -.700, \\ s_1 &= 3\ 090, & s_2 &= 1.619, \\ r &= -\ .6632, \end{aligned} \tag{17}$$

and $\bar{x}_1=170$, $\bar{x}_2=2.75$, and $\bar{x}_3=3\ 5$. The equations for predicting yield and purity in experiments are

$$
\begin{aligned}
Y_1 &= 58.529 - .3829(x_1 - 170) - 5.050(x_2 - 2.75) \\
&\qquad + 2.308(x_3 - 3.5),
\end{aligned} \tag{18}
$$

$$
\begin{aligned}
Y_2 &= 98.675 + .1558(x_1 - 170) + 4.144(x_2 - 2.75) \\
&\qquad - .700(x_3 - 3.5).
\end{aligned} \tag{19}
$$

The estimates of the coefficients of (4) are $b_0 = 301\,058$, $b = -2.4572$, $b_2 = 5.1338$ and $b_3 = .5883$ and the estimate of the variance of v is $s_1^2 - 2bs_1s_2r + b^2s_2^2 = 9.068$ (and standard deviation 3.011). The equation for predicting yield at a given purity is

$$
\begin{aligned}
Y_1 &= 301\,058 - 2.4572Y_2 + 5.1338(x_2 - 2.75) \\
&\qquad + .5883(x_3 - 3\;5) \\
&= 58.592 - 2.4572(Y_2 - 98.675) + 5.1338(x_2 - 2.75) \\
&\qquad + .5883(x_3 - 3\;5).
\end{aligned} \tag{20}
$$

6. CONFIDENCE INTERVALS AND TESTS OF HYPOTHESES

We base confidence intervals and tests of hypotheses on the theory that the pair of estimates (a_{1j}, a_{2j}) have a bivariate normal distribution with means $(\alpha_{1j}, \alpha_{2j})$ and variances $(\sigma_1^2/c_j, \sigma_2^2/c_j)$ and correlation ρ (i e , covariance $\sigma_1\sigma_2\rho/c_j$) Because of the orthogonality of the design (i e , $\sum_\gamma(x_{1\gamma} - \bar{x}_1)(x_{2\gamma} - \bar{x}_2) = 0$), one pair of estimates is independent of each other pair. If we want a confidence interval for a given coefficient α_{ij}, we use the usual student-t procedure; the confidence interval with confidence coefficient $1 - \epsilon$ consists of all α_{ij}^* satisfying

$$
\sqrt{c_j}\,\frac{|a_{ij} - \alpha_{ij}^*|}{s_i} \leqq t_{N-4}(\epsilon), \tag{21}
$$

where $t_{N-4}(\epsilon)$ is the two-tailed significance point of the t-distribution with $N-4$ degrees of freedom at significance level ϵ If we want a confidence region for a pair $(\alpha_{1j}, \alpha_{2j})$ we use the T^2 procedure; the confidence region with confidence coefficient $1 - \epsilon$ consists of all pairs α_{1j}^*, α_{2j}^* satisfying

$$
\begin{aligned}
&\frac{c_j}{1 - r^2}\left[\frac{(a_{1j} - \alpha_{1j}^*)^2}{s_1^2} - 2r\frac{(a_{1j} - \alpha_{1j}^*)(a_{2j} - \alpha_{2j}^*)}{s_1s_2} + \frac{(a_{2j} - \alpha_{2j}^*)^2}{s_2^2}\right] \\
&\qquad\qquad \leqq 2\,\frac{N-4}{N-5}\,F_{2,N-5}(\epsilon),
\end{aligned} \tag{22}
$$

where $F_{2,N-5}(\epsilon)$ is the upper significance point of the F-distribution with 2 and $N-5$ degrees of freedom at significance level ϵ. For example,

a confidence region with confidence coefficient .95 for the pair of coefficients of temperature is

$$\frac{[\alpha_{11}^* - (-.3829)]^2}{.01652} + 2\,\frac{[\alpha_{11}^* - (-.3829)][\alpha_{21}^* - .1558]}{.01305} + \frac{[\alpha_{21}^* - 1558]^2}{.004533} \leq 1 \tag{23}$$

To test the hypothesis that $\alpha_{11}=0$ and $\alpha_{21}=0$, we observe that $\alpha_{11}^*=0$ and $\alpha_{21}^*=0$ does not satisfy the above inequality, and hence reject this null hypothesis at the 05 significance level.

Now let us consider the coefficient β in (4). Any given linear combination $a_{11}-\beta^* a_{21}$ is normally distributed with mean $\alpha_{11}-\beta^*\alpha_{21}$ and variance $[\sigma_1^2-2\beta^*\rho\sigma_1\sigma_2+(\beta^*)^2\sigma_2^2]/c_1$. A test of the hypothesis that $\beta=\beta^*$ is equivalent to a test that $a_{11}-\beta^* a_{21}$ has mean zero Under this null hypothesis $a_{11}-\beta^* a_{21}$ is normally distributed with mean zero and independently of $s_1^2-2\beta^* s_1 s_2 r+(\beta^*)^2 s_2^2$. Application of the usual least squares theory to $y_1-\beta^* y_2$ shows that $(N-4)(s_1^2-2\beta^* s_1 s_2 r+(\beta^*)^2 s_2^2)/(\sigma_1^2-2\beta^*\sigma_1\sigma_2\zeta+(\beta^*)^2\sigma_2^2)$ has the χ^2-distribution with $N-4$ degrees of freedom Thus under the hypothesis $\sqrt{c_1}(a_{11}-\beta^* a_{21})/\sqrt{s_1^2-2\beta^* s_1 s_2 r+(\beta^*)^2 s_2^2}$ has the t distribution with $N-4$ degrees of freedom. The hypothesis is rejected if

$$\frac{c_1(a_{11} - \beta^* a_{21})^2}{s_1^2 - 2\beta^* s_1 s_2 r + (\beta^*)^2 s_2^2} \geq t_{N-4}^2(\epsilon). \tag{24}$$

A confidence interval for β consists of all β^* for which the corresponding null hypothesis is not rejected, that is, all β^* not satisfying the above inequality This interval has the end points

$$\frac{c_1 a_{11} a_{21} - t_{N-4}^2(\epsilon) s_1 s_2 r}{c_1 a_{21}^2 - t_{N-4}^2(\epsilon) s_2^2} \pm \frac{\sqrt{c_1 t_{N-4}^2(\epsilon)[a_{11}^2 s_2^2 - 2a_{11}a_{21}s_1 s_2 r + a_{21}^2 s_1^2] - t_{N-4}^2(\epsilon) s_1^2 s_2^2(1-r^2)}}{c_1 a_{21}^2 - t_{N-4}^2(\epsilon) s_2^2} \tag{25}$$

In our case the endpoints of a confidence interval with confidence .95 are -2.747 ± 962.

We are also interested in confidence intervals for the other coefficients in (4). Unfortunately, no method is known for obtaining an exact confidence interval for one of the other coefficients However, we can obtain a joint confidence interval for β and one other coefficient, say β_3. For any specified β^*, $a_{13}-\beta^* a_{23}$ is normally distributed with

mean $\alpha_{13}-\beta^*\alpha_{23}$ and variance $[\sigma_1^2-2\beta^*\sigma_1\sigma_2\rho+(\beta^*)^2\sigma_2^2]/c_3$ and independently of $a_{11}-\beta^*a_{21}$ (because the pair (a_{13}, a_{23}) is independent of the pair (a_{11}, a_{21})) A test of the hypothesis that $\beta=\beta^*$ and $\beta_3=\beta_3^*$ is equivalent to a test that $a_{11}-\beta^*a_{12}$ has mean zero and that $a_{13}-\beta^*a_{23}$ has mean $\alpha_{13}-\beta\alpha_{23}=\beta_3^*$. This hypothesis is rejected if

$$\frac{c_1(a_{11}-\beta^*a_{21})^2+c_3(a_{13}-\beta^*a_{23}-\beta_3^*)^2}{s_1^2-2\beta^*s_1s_2r+(\beta^*)^2s_2^2} \geqq 2F_{2,20}(\epsilon), \tag{26}$$

where $F_{2,20}(\epsilon)$ is the ϵ significance point of the F-distribution with 2 and 20 degrees of freedom A confidence region for the pair (β, β_3) consists of all pairs (β^*, β_3^*) which do not satisfy the above inequality. The confidence region can be written as

$$\begin{aligned}(c_1a_{21}^2+c_3a_{23}^2-2Fs_2^2)(\beta^*-m)^2+2c_3a_{23}(\beta^*-m)&\\ \cdot[\beta_3^*-(a_{13}-a_{23}m)]+c_3[\beta_3^*-(a_{13}-a_{23}m)]^2&\\ \leqq \frac{2Fc_1(a_{11}^2s_2^2+a_{21}^2s_1^2-2a_{11}a_{21}s_1s_2r)-4F^2s_1^2s_2^2(1-r^2)}{c_1a_{21}^2-2Fs_2^2}&,\end{aligned} \tag{27}$$

where $F=F_{2,20}(\epsilon)$ and

$$m=\frac{c_1a_{11}a_{21}-2Fs_1s_2r}{c_1a_{21}^2-2Fs_2^2} \tag{28}$$

In our case, for $1-\epsilon=.95$, we have the region

$$\begin{aligned}.8839[\beta^*-(-3\,002)]^2-1.0056[\beta^*-(-3.002)]&\\ [\beta_3^*-.207]+7183[\beta_3^*-.207]^2\leqq 1.&\end{aligned} \tag{29}$$

The center of the ellipse is at $\beta^*=-3.002$, $\beta_3^*=.207$. The maximum value of β^* in the ellipse is $-1\,641$ and the minimum, $-4\,363$, the maximum value of β_3^* is 1.728 and the minimum -1.314. It is seen that there are pairs (β^*, β_3^*) with $\beta_3^*=0$ included in the ellipse, thus, there are null hypotheses with $\beta_3=0$ which would be accepted by this test procedure, that is, the procedure would lead to acceptance of some null hypotheses specifying β and specifying that x_3 does not really enter (4).

The hypothesis that $\beta_3=0$ is equivalent to the hypothesis that $\alpha_{13}\alpha_{21}-\alpha_{11}\alpha_{23}=0$, that is, that the matrix

$$\begin{pmatrix}\alpha_{11} & \alpha_{13}\\ \alpha_{21} & \alpha_{23}\end{pmatrix} \tag{30}$$

is of rank one (rank zero being regarded as impossible here). Tests of

a hypothesis of rank usually involve the roots of a certain determinantal equation, in this case

$$(31)\quad \begin{vmatrix} c_1a_{11}^2+c_3a_{13}^2-\lambda(N-4)s_1^2 & c_1a_{11}a_{21}+c_3a_{13}a_{23}-\lambda(N-4)s_1s_2r \\ c_1a_{11}a_{21}+c_3a_{13}a_{23}-\lambda(N-4)s_1s_2r & c_1a_{21}^2+c_3a_{23}^2-\lambda(N-4)s_2^2 \end{vmatrix} = 0.$$

The roots are 3 352 and .058. The likelihood ratio criterion is $(1+.058)^{-(1/2)N}$ The hypothesis is to be rejected (on the basis of asymptotic theory) if $N \log_e (1+.058)$ exceeds the significance point of the χ^2-distribution with one degree of freedom [1] In this case the number $24 \log_e (1+ 058) = 1\,344$ which is less than the 5% or 10% significance point Alternative criteria to be compared with this significance point (such as $N\times$ 058 or $(N-4)\times .058$) are also small.

In the present problem, however, the fact that we cannot reject the null hypothesis of $\beta_3=0$ does not seem to be sufficient reason for eliminating x_3 from (4) The confidence region for β, β_3 admits values of β_3 as large as 1.728 If β_3 is actually this large it has engineering significance, that is, then the use of liquid B in the process does increase the yield of raffinate at the desired purity level It also has economic significance if the cost of using a certain flow of liquid B is less than the profit due to increased flow of raffinate. Since the statistical procedure does not rule out the economic significance, we do not eliminate x_3 from (4) It should be noted that if the confidence region included only very small values of β_3 (so small that admitted values could have no engineering or economic significance), we would be led to eliminate x_3 from (4)

A confidence region for all of the coefficients in (4) is given by

$$(32)\quad \frac{\begin{matrix} N(a_{10}-\beta^*a_{20}-\beta_0^*)^2 + c_1(a_{11}-\beta^*a_{21})^2 + c_2(a_{12}-\beta^*a_{22}-\beta_2^*)^2 \\ + c_3(a_{13}-\beta^*a_{23}-\beta_3^*)^2 \end{matrix}}{s_1^2 - 2\beta^*s_1s_2r + (\beta^*)^2s_2^2} \leqq 4F_{4,20}(\epsilon).$$

If the above expression is expanded, the confidence region can be expressed as

$$(33)\qquad \sum d_{ij}(\beta_i^* - e_i)(\beta_j^* - e_j) \leqq g,$$

where the sum runs over $i=0$, 2, 3 and blank (i e., $\beta_i^*=\beta^*$). If $c_1a_{21}^2/s_2^2>4F_{4\ 20}(\epsilon)$, then (d_{ij}) is a positive definite matrix and (33) is an ellipsoid in the space of β_0^*, β_2^*, β_3^*, and β^*. The generalized Schwartz inequality states

$$\begin{aligned}[\textstyle\sum h_i d_{ij}(\beta_j^* - e_j)]^2 &= [\textstyle\sum h_i d_{ij}\beta_j^* - \sum h_i d_{ij} e_j]^2 \\ &\leqq [\textstyle\sum d_{ij}(\beta_i^* - e_i)(\beta_j^* - e_j)][\sum d_{ij} h_i h_j]\end{aligned} \tag{34}$$

for any numbers h_i. Thus

$$\begin{aligned}g\sqrt{\textstyle\sum d_{ij}h_j} - \textstyle\sum h_i d_{ij} e_j &\leqq \textstyle\sum h_i d_{ij}\beta_j^* \\ &\leqq g\sqrt{\textstyle\sum d_{ij}h_i h_j} + \textstyle\sum h_i d_{ij} e_j\end{aligned} \tag{35}$$

for all h_i. This fact can be used to obtain confidence regions on $\mathcal{E}y_1$ for all possible choices of y_2, x_2, and x_3 in the production process, for

$$\mathcal{E}y_1 = \beta_0 + \beta y_2^* + \beta_2(x_2^* - \bar{x}_2) + \beta_3(x_3^* - \bar{x}_3), \tag{36}$$

when $y_2 = y_2^*$, $x_2 = x_2^*$, and $x_3 = x_3^*$. (36) is of the form $\sum h_i d_{ij}\beta_j$ for appropriate choices of h_i. The details of this procedure are too involved to include in this paper However, the procedure is important because it allows the investigator to make confidence statements about the expected yield in the production process for any choices of quality and flows of liquids A and B

APPENDIX. MORE GENERAL FORMULAS

In the experiment analyzed above the factors were applied at two levels and the design was balanced In this appendix we give some formulas for the more general case Suppose that the two regression equations suitable for the experimental data are

$$y_i = \alpha_{i0} + \sum_{j=1}^{p} \alpha_{ij}(x_j - \bar{x}_j) + u_i, \qquad i = 1, 2. \tag{37}$$

We assume N runs are made. Let

$$c_{jk} = \sum_{\gamma=1}^{N} (x_{j\gamma} - \bar{x}_j)(x_{k\gamma} - \bar{x}_k), \tag{38}$$

where $\bar{x}_j = \sum_\gamma x_{j\gamma}/N$. The estimate of α_{i0} is $a_{i0} = \bar{y}_i = \sum_\gamma y_{i\gamma}/N$ and the estimates of $\alpha_{i1}, \cdots, \alpha_{ip}$ are the solutions of

$$\sum_{k=1}^{p} c_{jk} a_{ik} = \sum_{\gamma=1}^{N} y_{i\gamma}(x_{j\gamma} - \bar{x}_j) \tag{39}$$

for $i = 1, 2$ The set of $2p$ estimates have in theory a joint normal distribution, the expected values of the estimates are the parameters and

$$\mathcal{E}(a_{ij} - \alpha_{ij})(a_{ik} - \alpha_{ik}) = c^{jk}\sigma_i^2, \tag{40}$$

$$\mathcal{E}(a_{1j} - \alpha_{1j})(a_{2k} - \alpha_{2k}) = c^{jk}\sigma_1\sigma_2\rho, \tag{41}$$

where the matrix (c^{jk}) is the inverse of (c_{jk}). The equation for predicting y_1 when x_1 is adjusted to make y_2 a given value is

$$y_1 = \beta_0 + \beta y_2 + \sum_{j=2}^{k} \beta_j(x_j - \bar{x}_j) + v, \tag{42}$$

where β_0, β, and v are given by (5), (6), and (9) and $\beta_j = \alpha_{1j} - \beta\alpha_{2j}$, $j=2, \cdots, p$, and the variance of v is given by (11) The estimates $b_0, b, b_2, \cdots, b_p$ are found from a_{ij} as $\beta_0, \beta, \beta_2, \cdots, \beta_p$ are found from α_{ij}. The estimates of σ_1^2 and σ_2^2 are found from

$$\begin{aligned}(N-p-1)s_i^2 &= \sum_{\gamma=1}^{N}\left[y_{i\gamma} - a_{i0} - \sum_{j=1}^{p} a_{ij}(x_{j\gamma} - \bar{x}_j)\right]^2 \\ &= \sum_{\gamma=1}^{N} y_{i\gamma}^2 - Na_{i0}^2 - \sum_{j,k=1}^{p} c_{jk}a_{ij}a_{ik},\end{aligned} \tag{43}$$

and r is found from

$$\begin{aligned}(N-p-1)rs_1s_2 &= \sum_{\gamma=1}^{N}\left[y_{1\gamma} - a_{10} - \sum_{j=1}^{p} a_{1j}(x_{j\gamma} - \bar{x}_j)\right] \\ &\qquad \left[y_{2\gamma} - a_{20} - \sum_{k=1}^{p} a_{2j}(x_{j\gamma} - \bar{x}_j)\right] \\ &= \sum_{\gamma=1}^{N} y_{1\gamma}y_{2\gamma} - Na_{10}a_{20} - \sum_{j,k=1}^{p} c_{jk}a_{1j}a_{2k}.\end{aligned} \tag{44}$$

The confidence interval for α_{ij} is

$$\frac{|a_{ij} - \alpha_{ij}^*|}{s_i c^{jj}} \leqq t_{N-p-1}(\epsilon). \tag{45}$$

The confidence region for (a_{1j}, a_{2j}) is given by (22) with c_j replaced by $1/c^{jj}$, with $(N-4)/(N-5)$ replaced by $(N-p-1)/(N-p-2)$ and with $F_{2,N-5}(\epsilon)$ replaced by $F_{2,N-p-2}(\epsilon)$ To test the hypothesis $\beta=\beta^*$ or to find a confidence region for β we use (24) and (25), respectively, with c_1 replaced by $1/c^{11}$ and $t_{N-4}(\epsilon)$ replaced by $t_{N-p-1}(\epsilon)$. To test a hypothesis $\beta=\beta^*$, $\beta_j=\beta_j^*$ we use the rejection region

$$\frac{\begin{aligned}c^{jj}(a_{11} - \beta^*a_{21})^2 - 2c^{1j}(a_{11} - \beta^*a_{21})(a_{1j} - \beta^*a_{2j} - \beta_j^*) \\ + c^{11}(a_{1j} - \beta^*a_{2j} - \beta_j^*)^2\end{aligned}}{(c^{11}c^{jj} - (c^{1j})^2)(s_1^2 - 2\beta^*s_1s_2r + (\beta^*)^2s_2^2)} \geqq 2F_{2,N-p-1}(\epsilon). \tag{46}$$

To test that $(\alpha_{11}\alpha_{2j}-\alpha_{21}\alpha_{1j})=0$, we use the roots of

$$\begin{vmatrix} D_{11} - \lambda(N-p-1)s_1^2 & D_{12} - \lambda(N-p-1)s_1s_2r \\ D_{21} - \lambda(N-p-1)s_1s_2r & D_{22} - \lambda(N-p-1)s_2^2 \end{vmatrix} = 0, \tag{47}$$

where

$$D_{hi} = \frac{a_{h1}a_{i1}c^{jj} - a_{h1}a_{ij}c^{1j} - a_{hj}a_{i1}c^{1j} + a_{hj}a_{ij}a^{11}}{c^{11}c^{jj} - (c^{1j})^2} \tag{48}$$

The derivations of these procedures are found in [1].

REFERENCES

[1] Anderson, T. W., "Estimating linear restrictions on regression coefficients for multivariate normal distributions," *Annals of Mathematical Statistics*, 22 (1951), 327–51

[2] Anderson, T W, and Rubin, Herman, "Estimation of the parameters of a single equation in a complete system of stochastic equations," *Annals of Mathematical Statistics*, 20 (1949), 46–63

[3] Berkson, J, "Are there two regressions?", *Journal of the American Statistical Association*, 45 (1950), 164–80.

[4] Haavelmo, T, "Methods of measuring the marginal propensity to consume," *Journal of the American Statistical Association*, 42 (1947), 105–22.

[5] Lindley, D V, "Estimation of a functional relationship," *Biometrika*, 40 (1953), 47–49.

SOME MINIMUM COST EXPERIMENTAL PROCEDURES IN QUADRATIC REGRESSION*

A. DE LA GARZA, D. S HAWXHURST, AND L T. NEWMAN
Carbide and Carbon Chemicals Company, Oak Ridge, Tennessee

Minimum cost procedures for multi-stage experiments determine the amount of replication at the various stages so that a desired estimate is obtained with a specified precision at minimum cost This paper extends these procedures to experiments in which the characteristic $\hat{y}$ being estimated is related to a controlled variable x by $\hat{y}=\alpha+\beta x+\gamma x^2$. Minimum cost procedures are obtained for estimating $\hat{y}$ with at least a specified precision over the permitted experimental range of x. The observations of $\hat{y}$ are assumed to be uncorrelated and to have known equal variances and costs independent of x Application is made to problems in chemical experimentation.

1 INTRODUCTION

THE problem of obtaining estimates with a specified precision from experiments at minimum cost is becoming increasingly important in industrial research and development In many of these experiments, costs and errors arise in well-defined stages For example, in estimating a chemical concentration of a batch, the batch is sampled and the samples are analyzed. Both of these stages are sources of cost and error Provided the costs and standard errors involved are known, minimum cost procedures for such experiments are easily obtained, see, for example, reference [3] These procedures determine the amount of replication required at the various stages so that the desired estimate may be obtained with the specified precision at minimum cost.

This paper considers minimum cost procedures for experiments in which the characteristic $\hat{y}$ being estimated is related to a controlled variable x by the quadratic relation $\hat{y}=\alpha+\beta x+\gamma x^2$ Estimates of the characteristic $\hat{y}$ with at least a specified precision are desired over the permissible experimental range of x at minimum cost It is seen that in these experiments, besides the usual question of the amount of replication, there is also the problem of determining values of the controlled variable x at which observations of $\hat{y}$ should be made A solution to this problem is given for uncorrelated observations of $\hat{y}$ with known constant variance and costs independent of x.

* Work done under AEC Contract No W-7405-eng-26.

2. OPTIMUM LOCATION

The minimum cost problems considered are illustrated by the following example:

Data are collected in a pilot plant to find the relation of the expected chemical concentration $\mathcal{Y}$ of pilot plant product with an operating pressure x. An observation of $\mathcal{Y}$ is the measured chemical concentration y of the product batch of a pilot plant run Due to equipment restrictions, the pilot plant must be operated in the pressure range x_L to x_H. After the data have been taken, a smooth curve is fitted to the observed points and an estimate of $\mathcal{Y}$ for any pressure in the range x_L to x_H may be obtained. It is desired that the precision of the estimates of $\mathcal{Y}$ shall not exceed a specified upper limit in the range x_L to x_H

In this experiment, the problems are to determine the required number N_s of observations y_i, $i=1(1)N_s$, and for what values of x they should be obtained so that the specified precision demand is satisfied at minimum cost To solve these problems for a reasonable number of situations, the following assumptions are made.

(1) the cost of an observation y_i is independent of x;
(2) the observations y_i are uncorrelated and have constant variance V,
(3) the x values do not have error;
(4) the expectation of an observation y_i taken at x is $E(y_i)=\mathcal{Y}$,
(5) the $\mathcal{Y}$ versus x relation is adequately represented by the quadratic

$$\mathcal{Y} = \alpha + \beta x + \gamma x^2;$$

(6) the estimates of $\mathcal{Y}$ are given by

$$Y = a + bx + cx^2,$$

where a, b, and c are the least squares estimates of α, β, and γ

Under the above assumptions (2) to (5), it is shown in reference [1] that if N observations y_i, $i=1(1)N$, are taken, the observations being limited to the range x_L to x_H, the spacing of these observations which minimizes the maximum variance of the Y values in the range x_L to x_H is $N/3$ observations at x_L, $(x_L+x_H)/2$, and x_H. The maximum variance of the Y values is then $3V/N$ Hence, if V_s is the variance corresponding to the specified precision limit on the Y values, the required number N_s of observations is determined from

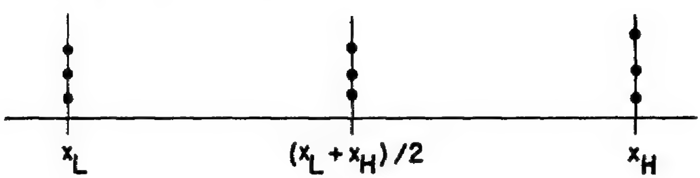
$$V_s = 3V/N_s. \tag{1}$$

For observations with cost independent of location x, the above number and spacing of observations constitutes a minimum cost procedure, inasmuch as the permitted maximum variance V_s of the Y values cannot be attained with fewer than N_s observations regardless of their spacing in the interval (x_L, x_H).

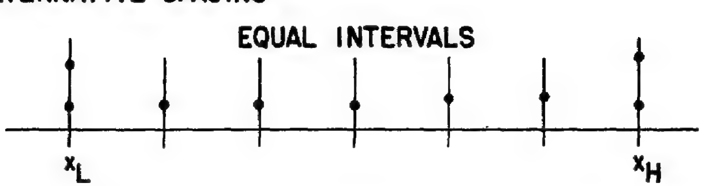
It may be found rather disturbing that observations are concentrated so heavily. In practice it is doubtful whether it is desirable to have observations at only three x's. It may be that the data are better fitted by a polynomial of higher degree than a quadratic, but as long as observations are located at three x's, this could never be detected. Fortunately, the spacing may be changed within reasonable limits from the minimum cost spacing without seriously affecting the precision of the Y values This is best shown by an example.

Assume that a minimum cost spacing requires nine observations; accordingly, three observations are located at x_L, $(x_L+x_H)/2$, and x_H. As an alternative, consider spacing the nine observations as follows: two observations at x_L, two observations at x_H, the remaining five observations being located at $x_L+k(x_H-x_L)/6$, $k=1(1)5$. These spacings are shown below.

MINIMUM COST SPACING



ALTERNATIVE SPACING



If in fact $\hat{y}=\alpha+\beta x+\gamma x^2$, the standard error of the Y values varies with x for both spacings as shown in Table I.

TABLE I

P	$F(x)$ for $N=9$	
	Minimum cost spacing	Alternative spacing
0.0	0 577	0 569
0.1	0 573	0.564
0 2	0 561	0.550
0.3	0 541	0 528
0.4	0.516	0 502
0.5	0.490	0.476
0.6	0 467	0.458
0.7	0.457	0 459
0 8	0 467	0 490
0.9	0 506	0 555
1 0	0 577	0.656

P is the deviation, positive or negative of x from the mid-point measured in standard units, thus:

$$P = \left| \left[x - \tfrac{1}{2}(x_H + x_L)\right] / \tfrac{1}{2}(x_H - x_L) \right|.$$

$F(x)$ is the relative standard error of the Y value at x, thus

$$F(x) = (\text{standard error of } Y \text{ value at } x)/\sqrt{\bar{V}}.$$

Table I clearly indicates that $F(x)$ for the alternative spacing does not depart greatly from $F(x)$ for the minimum cost spacing in case the $\hat{y}$ versus x relation is $\hat{y}=\alpha+\beta x+\gamma x^2$. If the $\hat{y}$ versus x relation is better fitted by a higher degree polynomial, the alternative spacing may supply evidence of this, while the minimum cost spacing cannot It should be pointed out, however, that the spacing of two observations at each end point, the others being distributed at equal intervals between the end points, in general is not to be recommended. For nine observations, such a spacing compares favorably with the corresponding minimum cost spacing as shown in Table I, but the comparison becomes progressively less favorable when the number of observations is increased.

In conclusion, a procedure which should be satisfactory in most situations is to space the observations at more than three locations so that there will be little increase in the maximum standard error of the Y values as compared to the maximum standard error given by the minimum cost spacing With this procedure, protection is provided against failure of the quadratic hypothesis, and if the hypothesis is true, little is lost in the precision of the estimates of $\hat{y}$.

3. OPTIMUM LOCATION AND AMOUNT OF REPLICATION

The experiments considered here are an extension of those in Part 2. These problems concern both the location of observations and the

replication of experimental stages. For illustration, the pilot plant setup of Part 2 will be followed.

Suppose that in the pilot plant experiment the taking of an observation y_i consists of adjusting operating conditions to the pressure x, waiting for equilibrium, taking samples of the product batch for the run, and having these samples analyzed by the laboratory for the desired chemical concentration. Thus, the variance of an observation y_i may be due to small fluctuations in pilot plant operation, some sampling error, and laboratory analytical error. Suppose further that n_1 laboratory analyses per sample are made and n_2 samples per product batch are taken The variance V of an observation then is

$$V = V_3 + V_2/n_2 + V_1/n_2n_1, \tag{2}$$

where: V_1 is the variance of a laboratory analysis, V_2 is the variance of a sample, and V_3 is the variance due to pilot plant fluctuations during a run producing a product batch It is assumed that V_1, V_2, and V_3 are known On the cost side, suppose that the costs are of the form f_i+u_in, where f_i is a fixed cost and u_i is an incremental unit cost For example, the cost of n laboratory analyses would be f_1+u_1n, and similarly for samples, with a subscript 2, and for runs, with a subscript 3 As before, it is demanded of the experiment that the precision of the Y values shall at no place exceed a specified limit in the range x_L to x_H. The cost problem is to determine the number of product batches, the number of samples per product batch, and the number of analyses per sample so that the specified precision demand is satisfied with minimum cost

The solution to this problem is not at all difficult Let V_s be the variance corresponding to the specified precision limit of the Y values. Since the variance of the observations y_i is constant, it follows from (1) and (2) that

$$V_s = 3(V_3 + V_2/n_2 + V_1/n_2n_1)/N_s, \tag{3}$$

where N_s is the required number of observations y_i, $i=1(1)N_s$ Also, as before, the spacing of these N_s observations in the interval x_L to x_H is $N_s/3$ observations at x_L, $(x_L+x_H)/2$, and x_H Furthermore, the total cost of the experiment is

$$C_t = f_3 + f_2 + f_1 + u_3N_s + u_2N_sn_2 + u_1N_sn_2n_1. \tag{4}$$

The cost problem then is to find n_1, n_2, and N_s that minimize C_t subject to satisfying the expression for V_s. Using the technique of Lagrange multipliers, the answer is readily found to be

$$n_1 = (V_1u_2/V_2u_1)^{1/2}; \qquad n_2 = (V_2u_3/V_3u_2)^{1/2};$$
$$N_s = 3V_3^{1/2}[(V_3u_3)^{1/2} + (V_2u_2)^{1/2} + (V_1u_1)^{1/2}]/V_su_3^{1/2}. \tag{5}$$

The minimum total cost is found to be

$$\min C_t = f_3 + f_2 + f_1 + 3[(V_3u_3)^{1/2} + (V_2u_2)^{1/2} + (V_1u_1)^{1/2}]^2/V_s. \tag{6}$$

It is clear that here, as in Part 2, the N_s observations may be spaced at more than three x's at the expense of a small increase in the maximum standard error of the Y values.

4 DISCUSSION

Several comments applicable to these minimum cost problems are in order Possibly the point most needing discussion is the assumption made throughout that the variances of the experimental stages are known Unfortunately, it is not easy to determine variances With independent normal observations, if s^2 is an estimate of V with ν degrees of freedom, s^2/V is distributed like χ^2/ν. A check with Hald's tables, [2], shows, for example, that for $\nu=30$, the probability that $0.56 \leq s^2/V \leq 1.57$ is 95% It is thus apparent that the minimum cost procedures here given are strictly applicable only when well-established estimates of population variances are available At other times, only a rough preferred region is outlined. However, it is to be noted that even if not all the variances required are well-established, partial use of a minimum cost procedure is still possible. For example, it is seen in Part 3 that the required number of analyses per sample depends solely on the costs and standard errors of analyses and samples, this number of analyses may be determined even though the variance of a pilot plant run is not known

As in many optimization studies, once having determined the requirements at a minimum cost point, it is well to map the behavior of the cost in the neighborhood of the minimum. It may well happen that at a slightly greater cost, the requirements may be met with greater convenience not easily expressed in dollars and cents For example, a minimum cost point may require a number of analyses that overloads the laboratory capacity on that particular type of analysis and results in long time delays In such a case, the cost may be investigated around the minimum in a direction of reducing the number of analyses. Another reason for such investigation is that the variables considered are whole numbers When the numbers are small and the associated costs are high, rounding off an optimization to whole numbers should be done carefully.

It is of interest to note that expressions for minimum total cost procedures, such as (5) and (6), clearly point out how each factor in the experiment contributes to the total cost. Thus, it is possible to isolate the major factors contributing to the total cost, and efforts to reduce the costs and/or associated standard errors still further can be concentrated on the controlling factors.

In practice, costs are likely to have a linear form such as $f+un$ used in Part 3. Even if the costs are given by some more complicated form, approximation by $f+cn$ in the range of interest should be satisfactory. Variations of how such linear costs apply can be made easily. For example, in Part 2, let the cost of the $N_s n_2 n_1$ analyses also depend on the $N_s n_2$ samples, so that the cost of the analyses becomes $f_1 N_s n_2 + u_1 N_s n_2 n_1$. Such a cost for analyses could result from a preliminary laboratory treatment of each sample prior to running the analytical determinations. The total cost for the experiment would be

$$f_3 + f_2 + u_3 N_s + u_2 N_s n_2 + f_1 N_s n_2 + u_1 N_s n_2 n_1,$$

or

$$f_3 + f_2 + u_3 N_s + (u_2 + f_1) N_s n_2 + u_1 N_s n_2 n_1$$

A comparison of this with the total cost given by (4) shows readily how the cost minimization proceeds.

It is to be emphasized that the minimum cost procedures given here assume that the errors in measurements are not correlated. To say the least, this implies that the errors in measurements do not depend on observer, analytical instrument, etc. For example, some observers and some analytical instruments tend to have a "memory." In such situations, replication of measurements to decrease error is often completely useless

Finally, it may be seen that there are many situations for which the results of Part 2 may be coupled with expressions for costs and standard errors of experimental stages to obtain minimum cost procedures.

REFERENCES

[1] de la Garza, A., "Spacing of information in polynomial regression," *Annals of Mathematical Statistics*, 25 (1954), 123–30.

[2] Hald, A., *Statistical Tables and Formulas*, New York, John Wiley and Sons, Inc (1952).

[3] Hald, A., *Statistical Theory with Engineering Applications*, New York, John Wiley and Sons, Inc. (1952), 496–98.

TABLES OF PERCENTAGE POINTS FOR THE STUDENTIZED MAXIMUM ABSOLUTE DEVIATE IN NORMAL SAMPLES*

MAX HALPERIN, SAMUEL W. GREENHOUSE, JEROME CORNFIELD, AND JULIA ZALOKAR

National Institutes of Health

Tables of upper and lower limits on the upper 5 per cent and 1 per cent points of the distribution of the studentized maximum absolute deviate in normal samples are presented. The method of computation and the reliability of the tables are described, and approximations which may be used to supplement the table are derived and discussed Examples of the use of the tables are given with special attention devoted to their use for multiple significance testing on a set of means.

1. INTRODUCTION

LET $x_1, x_2, \cdots, x_k$ be independent normally distributed variates, each with mean μ and variance σ^2 Define the studentized maximum absolute deviate by

$$d = \max_{i=1,2,\cdots,k} \frac{|x_i - \bar{x}|}{s},$$

where ms^2/σ^2 is distributed as χ^2 with m degrees of freedom (d f) and independent of $x_1, x_2, \cdots, x_k$

In this paper we present tables of upper and lower limits on the upper 5 per cent and 1 per cent points of the distribution of d for varying d f. The statistic d is the two-sided version of the studentized extreme deviate of Nair [7]. Thus if we denote by $y_1, y_2, \cdots, y_k$ the ordered values of $x_1, x_2, \cdots, x_k$,

$$d = \max\left[\frac{y_k - \bar{y}}{s}, \frac{\bar{y} - y_1}{s}\right],$$

where the statistics in the square brackets are the two possible extreme deviates in a sample of size k so that each has the distribution tabled by Nair. Both d and the Nair statistic are to be contrasted with the extreme deviate statistic of Thompson [10] which differs fro of Nair in that the scaling statistic, s, is the sample stand and is thus not independent of the numerator.

* The authors wish to acknowledge the valuable computing assista

The tables given here may be used as the basis for a test of whether the largest observation, without regard to sign, is too large (i e as an outlier test, as may the related Nair and Thompson statistics) The tables may also be used as a basis for multiple significance tests of a set of k means arising from independent samples on normal populations with the same variances but possibly different means. The use of the tables as in an outlier test is illustrated by the first example of Section 2. In our second example, use of the tables in multiple testing of means is illustrated when each mean arises from the same size sample Our third example illustrates a discussion of the modification of computations and use of the tables for multiple testing of means when sample sizes are unequal. In Section 3, results necessary for computation of the tables are derived, and in Section 4 some approximate procedures are given for obtaining approximate significance points.

2. DESCRIPTION AND USE OF TABLES

Upper and lower limits to the upper 5 per cent and 1 per cent points of the distribution of d are given in Tables 1 and 2 for varying numbers of variates (k) and degrees of freedom (m) The correct critical value lies between the tabulated upper and lower limits From the derivation given in Section 3 it is clear that the lower limit is a closer approximation to the correct value than the upper, although the difference between the two will not often be of practical importance [1]

The selection of m as $\geq k$ arose from a consideration of the problem of multiple significance tests on a set of sample means, rather than the detection of outliers. From this point of view, one would expect at least 1 d f. for error from each sample In any event, approximate significance points can be obtained for $m<k$ by the procedures of Section 4

Example 1 Our first example is taken from Snedecor's *Statistical Methods* [9] and is the same as example 1 of Nair [7]

A randomized block experiment with four strains of wheat and five replications gives the following mean yield in pounds per plot.

A	B	C	D
34.4	34 8	33 7	28 4

The error variance is estimated as 2 19 with 12 d.f. Following Nair we compute $(\bar{x}-x_1)/s$, where s is the estimated standard error of any one

[1] Thus, the difference between 6.97 and 6 08, the upper and lower limits for $k=m=3$, $p=01$ corresponds to a difference in probability of approximately .003

of the means as $(32.8-28.4)/\sqrt{2\ 19/5}$ or 6.7, with $k=4$ and $m=12$. Reference to Table 2 shows this deviation to be significant at less than the 01 level, since 6.7 far exceeds the upper limit for $m=10$, which is, of course, greater than that for $m=12$.

TABLE 1

UPPER AND LOWER LIMITS FOR 5% POINT OF DISTRIBUTION OF d

(k = sample size, m = d.f.)

m/k	3	4	5	6	7	8	9	10	15	20	30	40	60
3	3 97												
	3 35												
4	3 24	3 74											
	2 89	3 27											
5	2 89	3 30	3 61										
	2 65	3 01	3 24										
6	2 68	3 05	3 32	3 53									
	2 50	2 84	3 06	3 22									
7	2 55	2 89	3 13	3 32	3 48								
	2 40	2 73	2 94	3 10	3 22								
8	2 46	2 78	3 00	3 18	3 32	3 44							
	2 33	2 64	2 85	3 01	3 13	3 22							
9	2 40	2 70	2 91	3 07	3 21	3 32	3 41						
	2 28	2 58	2 78	2 94	3 06	3 15	3 21						
10	2 34	2 63	2 84	2 99	3 12	3 23	3 32	3 40					
	2 24	2 53	2 73	2 88	2 99	3 08	3 15	3 23					
15	2 20	2 46	2 64	2 77	2 88	2 97	3 05	3 12	3 37				
	2 12	2 39	2 57	2 71	2 81	2 90	2 97	3 03	3 27				
20	2 13	2 38	2 55	2 67	2 77	2 86	2 93	2 99	3 22	3 38			
	2 07	2 32	2 50	2 62	2 73	2 81	2 88	2 94	3 15	3 29			
30	2 07	2 30	2 46	2 58	2 67	2 75	2 82	2 88	3 08	3 22	3 40		
	2 01	2 26	2 42	2 54	2 64	2 72	2 78	2 84	3 04	3 17	3 34		
40	2 04	2 27	2 42	2 54	2 63	2 70	2 77	2 82	3 02	3 15	3 32	3 43	
	1 99	2 23	2 39	2 50	2 60	2 68	2 73	2 79	2 99	3 11	3 27	3 39	
60	2 01	2 23	2 38	2 49	2 58	2 65	2 71	2 77	2 96	3 08	3 24	3 35	3 49
	1 96	2 20	2 35	2 47	2 56	2 63	2 69	2 75	2 93	3 05	3 21	3 32	3 45
120	1 98	2 20	2 34	2 45	2 54	2 61	2 66	2 72	2 89	3 01	3 16	3 27	3 40
	1 93	2 17	2 32	2 43	2 52	2 59	2 64	2 70	2 88	2 99	3 14	3 25	3 38
∞	1 95	2 16	2 30	2 41	2 49	2 56	2 61	2 66	2 83	2 95	3 09	3 19	3 31
	1 91	2 14	2 28	2.39	2 48	2 55	2 60	2 65	2 82	2 94	3 08	3 18	3 30

TABLE 2

UPPER AND LOWER LIMITS FOR 1% POINT OF DISTRIBUTION OF d

(k = sample size, m = d f)

m/k	3	4	5	6	7	8	9	10	15	20	30	40	60
3	6.97 6.08												
4	5.10 4.65	5.85 5.27											
5	4.28 4.03	4.86 4.54	5.27 4.87										
6	3.84 3.65	4.32 4.10	4.66 4.41	4.93 4.64									
7	3.56 3.42	3.98 3.83	4.28 4.13	4.52 4.33	4.71 4.49								
8	3.37 3.26	3.76 3.64	4.03 3.91	4.24 4.10	4.42 4.25	4.56 4.38							
9	3.23 3.14	3.59 3.50	3.85 3.76	4.04 3.94	4.20 4.08	4.32 4.21	4.44 4.30						
10	3.12 3.04	3.47 3.40	3.71 3.63	3.90 3.82	4.05 3.94	4.16 4.07	4.27 4.15	4.35 4.23					
15	2.84 2.79	3.14 3.10	3.34 3.30	3.49 3.45	3.61 3.57	3.71 3.67	3.79 3.75	3.87 3.81	4.15 4.07				
20	2.72 2.68	3.00 2.97	3.18 3.16	3.32 3.29	3.43 3.40	3.51 3.49	3.59 3.56	3.65 3.62	3.89 3.85	4.05 4.00			
30	2.60 2.57	2.86 2.84	3.03 3.01	3.16 3.14	3.26 3.24	3.34 3.32	3.40 3.38	3.46 3.44	3.67 3.65	3.80 3.78	3.98 3.95		
40	2.55 2.52	2.80 2.78	2.96 2.94	3.08 3.07	3.18 3.16	3.25 3.24	3.31 3.30	3.37 3.36	3.57 3.56	3.69 3.68	3.85 3.83	3.97 3.95	
60	2.50 2.47	2.74 2.72	2.89 2.88	3.01 2.99	3.10 3.09	3.17 3.16	3.23 3.22	3.28 3.27	3.47 3.46	3.59 3.58	3.74 3.73	3.85 3.83	3.98 3.97
120	2.45 2.42	2.68 2.66	2.83 2.82	2.94 2.93	3.03 3.02	3.10 3.09	3.15 3.14	3.20 3.20	3.38 3.37	3.49 3.48	3.62 3.62	3.73 3.71	3.87 3.86
∞	2.40 2.38	2.62 2.61	2.76 2.76	2.87 2.87	2.95 2.95	3.02 3.02	3.07 3.07	3.12 3.12	3.29 3.29	3.39 3.39	3.53 3.53	3.62 3.62	3.73 3.73

Example 2 Our second example, from Duncan [2], illustrates the kind of multiple statements that can be made concerning a set of means using the statistic d Given the 11 means, 2.85, 3.60, 3.94, 4.05, 4.55, 4.96, 5.03, 5.41, 5.96, 6.09, 6.97 each with standard error 0.34 (esti-

mated with 30 degrees of freedom) we compute the ratio of each of the 11 deviations from the over-all average to the standard error of the mean as −5.91, −3.70, −2 70, −2.37, −0 90, +0.31, +0.51, +1.63, +3.25, +3.64, +6.23. Table 1 shows that for $k=10$ and 15, $m=30$, the .05 critical value is no greater than 3 08 and no less than 2 83 The largest mean, 6.97, is consequently significantly above the mean of all 11 and may be considered an upward outlier. Similarly, the lowest mean, 2 85, is significantly below the mean of all 11 and may be considered a downward outlier. Furthermore, the two means, 5.96 and 6.09, are also significant upward outliers and the mean 3 60 is a significant downward outlier. We thus assert that each of the two lowest means is below the over-all average and each of the three highest is above the over-all average. The probability that this statement is incorrect when the null hypothesis is true is obviously .05 When the null hypothesis is false this probability is less than 05 (Appendix).

Remark on unequal numbers of cases in comparison of means

In the frequently occurring case in which the k means are based upon unequal numbers of cases the present tables may still be used, but with three modifications (*a*) The deviation is taken from the weighted over-all mean [2] (*b*) The ith deviation is divided, not by the standard error of the ith mean, but by this standard error times

$$\sqrt{\frac{(n-n_i)}{n}\,\frac{k}{(k-1)}} \quad \text{where} \quad n=\sum_{i=1}^{k} n_i.$$

(*c*) The tabulated upper limit is still an upper limit to the absolute maximum deviate so computed, but the lower limit may not be, as is obvious from the derivations in Section 3 In fact, the lower limit as we have computed it would vary with each configuration of n_i and cannot be practicably tabulated

Example 3 (from Snedecor [9]). We reproduce below the average dressing percentage (less 70 per cent) for five breeds of swine. The variance of a single observation, 5 51, is estimated with 83 degrees of freedom

---

[2] For problems in which the deviation from the unweighted mean is more suitable this may be used but the divisor in step (*b*) becomes

$$\left\{\frac{s^2}{k(k-1)}\left[\frac{k(k-2)}{n_i}+\sum_{j=1}^{k}\frac{1}{n_j}\right]\right\}^{1/2}$$

i Breed	Dressing %	n_i Number of observations	$\sqrt{5.51\left(\frac{93-n_i}{93 n_i}\right)\frac{5}{4}}$	$\dfrac{\lvert x_i-\bar{x}\rvert}{s_i\sqrt{\frac{93-n_i}{93}\cdot\frac{5}{4}}}$
1	14 25	18	556	3 40
2	12 12	45	281	0 85
3	11 48	6	1 036	0.85
4	12 53	9	832	0.20
5	11 05	15	621	2 11
	12 36	93		

Breed 1 gives a significantly higher than average dressing per cent at the .01 level

3 DERIVATION OF RESULTS AND COMPUTATION OF TABLES

We consider in detail the special case of σ known The case for σ estimated will be seen to be completely analogous. Thus let

$$z_i = \frac{x_i - \bar{x}}{\sigma} \qquad i = 1, 2, \cdots, k.$$

We then have,

$$Ez_i = 0,$$

$$\text{Var } z_i = \frac{k-1}{k}, \qquad i = 1, 2, \cdots, k$$

$$\text{Cov } z_i z_j = -\frac{1}{k}, \qquad i \neq j$$

Thus, $z_1, \cdots, z_k$ have a k-variate joint normal distribution of rank $k-1$

Now, for any significance level α, we define the positive number, h, by

$$\text{(3.1)} \qquad \Pr\left\{ |z_1| \leqq h; \cdots, |z_k| \leqq h \right\} = 1 - \alpha$$

It is obvious that (3 1) is equivalent to

$$\text{(3.2)} \qquad \Pr\left\{ \max |z_i| \leqq h \right\} = 1 - \alpha.$$

Hence a solution of (3 1) for h, or an approximation thereto, will yield a procedure for obtaining percentage points for $\delta = \lim_{m\to\infty} d$.

It is easy to obtain approximate solutions to (3 1) by taking advantage of the fact that the correlations between pairs of the z's is

$-1/(k-1)$, coupled with elementary probability considerations Specifically define $\bar{A}_i$ as the event $|z_i| > h$, $i=1, 2, \cdots, k$ and A_i as the event $|z_i| \leqq h$ We then have immediately for (3 1) by the inequalities of Bonferroni [4],

$$
\begin{aligned}
(3.3) \qquad 1 - \sum_{i=1}^{k} \Pr\{\bar{A}_i\} &\leqq \Pr\{A_1 A_2 \cdots A_k\} \\
&\leqq 1 - \sum_{i=1}^{k} \Pr\{\bar{A}_i\} + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \Pr\{\bar{A}_i \bar{A}_j\}.
\end{aligned}
$$

Then, since each z_i has a normal distribution with zero mean and variance $(k-1)/k$, and each pair, z_i, z_j, has a bivariate normal distribution with zero means, common variance $(k-1)/k$, and covariance $-1/k$, we can write

$$(3.4a) \quad \Pr\{\bar{A}_i\} = \frac{2}{\sqrt{2\pi}} \int_{\sqrt{(k/k-1)}\,h}^{\infty} \exp - \frac{x^2}{2}\, dx, \qquad i = 1, 2, \cdots, k,$$

and

$$
\begin{aligned}
(3.4b) \quad \Pr\{\bar{A}_i \bar{A}_j\} &= 2 \int_{\sqrt{(k/k-1)}\,h}^{\infty} \int_{\sqrt{(k/k-1)}\,h}^{\infty} p(x_1\, x_2)\, dx_1 dx_2 \\
&\quad + 2 \int_{-\infty}^{-\sqrt{(k/k-1)}\,h} \int_{\sqrt{(k/k-1)}\,h}^{\infty} p(x_1, x_2)\, dx_1 dx_2
\end{aligned}
$$

$i \neq j$, $i, j = 1, 2, \cdots, k$, where $p(x_1, x_2)$ is a bivariate normal density with the variates involved having zero means, unit variances, and correlation $-1/(k-1)$

Thus, (3 3) is the basis of our computation procedure. By equating the left hand side of (3.3) to $(1-\alpha)$ and solving for h, we clearly get an upper limit for h, h_U say Similarly equating the right hand side of (3.3) to $(1-\alpha)$ and solving for h, we get a lower limit for h, h_L say.

To go from the case of σ known to a σ estimated by say, s, where ms^2/σ^2 is a chi-square variate with m d f and independent of x_1, x_2, $\cdots$, x_k, we need only note that

$$
\begin{aligned}
&\Pr\left\{\left|\frac{x_1 - \bar{x}}{s}\right| \leqq h; \cdots; \left|\frac{x_k - \bar{x}}{s}\right| \leqq h\right\} \\
&\qquad = \Pr\left\{\left|\frac{x_1 - \bar{x}}{\sigma}\right| \leqq \frac{s}{\sigma} h; \cdots; \left|\frac{x_k - \bar{x}}{\sigma}\right| \leqq \frac{s}{\sigma} h\right\}.
\end{aligned}
$$

Thus the argument for the case of σ known holds with the single exception that h is everywhere replaced by $(s/\sigma)h$ so that we must

integrate over the distribution of s The result is that analogous to (3.4) we get

$$\Pr\{\overline{A}_i\} = \int_{\sqrt{(k/k-1)}\,h}^{\infty} p(t)dt, \qquad i = 1, 2, \cdots, k, \tag{3.5a}$$

and

$$\begin{aligned}\Pr\{\overline{A}_i\overline{A}_j\} &= 2\int_{\sqrt{(k/k-1)}\,h}^{\infty}\int_{\sqrt{(k/k-1)}\,h}^{\infty} p(t_1, t_2)dt_1dt_2 \\ &\quad + 2\int_{-\infty}^{-\sqrt{(k/k-1)}\,h}\int_{\sqrt{(k/k-1)}\,h}^{\infty} p(t_1, t_2)dt_1dt_2\end{aligned} \tag{3.5b}$$

$i \neq j$, $i, j = 1, 2, \cdots, k$, where $p(t)$ is the usual Student's t density with m d.f. and

$$p(t_1, t_2) = \frac{1}{2\pi\sqrt{1-\rho^2}}\left[1 + \frac{t_1^2 - 2\rho t_1t_2 + t_2^2}{m(1-\rho^2)}\right]^{-(m/2)-1}, \tag{3.6}$$

with $\rho = -1/(k-1)$ The density given by (3 6) is also derived in [3]

For the case of σ known, h_U was computed using the Bureau of Standards tables [8] of the normal integral, for σ estimated, h_U was computed using the t table of Hartley and Pearson [5] and the Pearson Tables of the Incomplete Beta Function [1] along with La Grangian 3 point interpolation where necessary. These solutions, of course, are practically immediate

To obtain lower limits for h, trial values of h were guessed for each k and m and the value of the right hand side of (3 3) was computed. Except for integrals of the bivariate density (3 6) this involved only computations similar to those for finding the upper limit for h. For the integral involving (3.6) the transformation

$$\begin{aligned} t_1 &= \frac{1}{\sqrt{2}}\left(\frac{r\sin\theta}{\sqrt{1-\rho}} + \frac{r\cos\theta}{\sqrt{1+\rho}}\right), \\ t_2 &= \frac{1}{\sqrt{2}}\left(\frac{r\sin\theta}{\sqrt{1-\rho}} - \frac{r\cos\theta}{\sqrt{1+\rho}}\right), \end{aligned} \tag{3.7}$$

enables us to reduce it to

$$1 - \frac{2}{\pi}\int_{-\pi/2+\arctan\sqrt{(1-\rho)(1+\rho)}}^{\arctan\sqrt{(1-\rho)(1+\rho)}} \left(1 + \frac{h^2\sec^2\theta}{m}\right)^{-m/2} d\theta, \tag{3.8}$$

which was evaluated by quadrature using Simpson's rule with 8 inter-

vals.[3] Successive trials were made using trial values of h_L to two decimal places until the value $(1-\alpha)$ was bracketed by two trial values differing by .01. Linear interpolation was then employed to decide which of these two values would be tabled. Thus the lower limits given in Tables 1 and 2 are correct to two decimal places, with a few exceptions. The exceptions are due to limitations of the interpolation procedure in the tables [1]. Numerical checks indicated that for cases where [1] was used, lower limits are too low, but, even then, by at most, .01. The upper limits (h_U) in both tables are correct to two decimal places.

The difficulties arising in obtaining significance points if we wish to use a maximum absolute deviate test in comparison of means based on unequal numbers of observations may be seen by considering the case when σ^2 is known and equal to one, say

We suppose $\bar{x}_1, \bar{x}_2, \cdots, \bar{x}_k$, to be means of independent samples of size $n_1, \cdots, n_k(\sum_{i=1}^{k} n_i = n)$ on a normal population with zero mean and unit variance. If we define

$$w_i = \frac{n_i}{n} \qquad i = 1, 2, \cdots, k$$

$$\bar{x} = \sum_{i=1}^{k} w_i \bar{x}_i$$

$$z_i = \frac{(\bar{x}_i - \bar{x})\sqrt{n_i}}{\sqrt{\dfrac{n - n_i}{n}\,\dfrac{k}{k-1}}} \qquad i = 1, 2, \cdots, k$$

easy computation shows that $\operatorname{Var} z_i = (k-1)/k$, $i = 1, 2, \cdots, k$ However we find that

$$\operatorname{cov} z_i z_j = -\frac{(k-1)}{k}\sqrt{\frac{n_i n_j}{(n-n_i)(n-n_j)}}, \qquad i \neq j.$$

Thus though (3 3) still holds for this case, it becomes impractical, as earlier remarked, to compute lower limits for h, since for each i and j, the bivariate integral could in the extreme case be different.

4. AN ALTERNATIVE PROCEDURE FOR OBTAINING APPROXIMATE SIGNIFICANCE POINTS

The inequality (3 3) is, of course, valid for any $\rho \geq -1/(k-1)$ [4] If we

[3] For infinite d f. (σ known) the limit of (3 8) was quadratured as described. It is easy to show that the limiting integral may also be obtained by transformation of the corresponding integral of the bivariate normal

[4] For $\rho < -1/(k-1)$, it is easily shown that the covariance matrix of $z_1, \cdots, z_k$ becomes indefinite.

investigate the behavior of the right hand side of (3.3) for variation in ρ, it is not difficult to show that the expression is monotone increasing in ρ^2 (whether σ^2 is known or estimated) It immediately follows that a lower limit, $h_L{}^*$, computed from (3 3) assuming $\rho=0$, will fall above the lower limit, h_L, computed for $\rho=-1/(k-1)$ Thus an approximate lower limit (for σ^2 known) can be computed by equating the right hand side of (3 3), with ρ assumed equal to zero, to $(1-\alpha)$ and solving for h. This gives

$$(4.1) \qquad \Pr\{\overline{A}_i\} = \frac{1}{k-1}\left[1 - \sqrt{1 - 2\frac{(k-1)}{k}\alpha}\right],$$

where $\Pr\{\overline{A}_i\}$ is defined by (3 4a) If σ^2 is estimated with m d f we can utilize a theorem proved by Halperin in [6] to show that (4 1) is still an approximate lower limit with $\Pr\{\overline{A}_i\}$ defined by (3 5a)

Also since (3 3) is valid for $\rho=0$, we can, from the above discussion, assert that h computed under the assumption of independence will lie between h_U and h_L and thus will serve as an approximate significance point That is, since for this case

$$(4.2) \qquad \begin{aligned} \Pr\{A_1 \cdots A_k\} &= [\Pr\{A_i\}]^k = 1-\alpha, \quad \text{we may use} \\ \Pr\{\overline{A}_i\} &= 1-(1-\alpha)^{1/k}, \end{aligned}$$

where $\Pr\{\overline{A}_i\}$ is defined by (3 5a) to compute approximate significance points

Some numerical investigation was carried out to explore the closeness of the approximation (4 2) It was found in all cases that the h obtained from (4 2) was only slightly below the tabulated upper limit Thus even for the extreme case $m=k=3$, the h given by (4 2) is 3 94 for $p=.05$ and 6.97 for $p=01$ as compared with 3 97 and 6 97 in Tables 1 and 2 Equation (4 1) was not investigated numerically, but clearly will yield results close to the lower limit h_L

Thus, it is suggested that (4 1) and (4 2) be used to obtain approximations to critical values for m k, and p not tabulated.

APPENDIX

In connection with example 2 of Section 2 we wish to show that the probability of our composite assertion being wrong in any respect is at most α whether or not the null hypothesis is true Such is, in fact, the case for more general situations than have been considered here and may be proved, almost trivially, as follows

Let $x_1, x_2, \cdots, x_k$ be a set of variates with means $\mu_1, \mu_2, \cdots, \mu_k$, unit variances, and arbitrary correlation matrix. Define h by

$$\text{(A.1)} \qquad \Pr \left\{ \begin{array}{c} -h \leqq (x_i - \mu_i) \leqq h \\ i = 1, 2, \cdots, k \end{array} \right\} = 1 - \alpha$$

Now let the first r of the k means be positive, the next $s-r$ negative, and the remaining $k-s$ zero (A.1) may then be rewritten as

$$\text{(A.2)} \quad \Pr \left\{ \begin{array}{lll} -h \leqq x_i - \mu_i \leqq h; & -h \leqq x_j - \mu_j \leqq h; & h \leqq x_l \leqq h \\ i = 1, 2, \cdots, r & j = r+1, \cdots, s & l = s+1, \cdots, k \end{array} \right\} = 1 - \alpha.$$

On the other hand the total probability of avoiding error is

$$\text{(A.3)} \quad \Pr \left\{ \begin{array}{ll} -h - \mu_i \leqq x_i - \mu_i \leqq \infty; & -\infty \leqq x_j - \mu_j \leqq h - \mu_j; \\ i = 1, 2, \cdots, r & j = r + 1, \cdots, s \\ & -h \leqq x_l \leqq h \\ & l = s + 1, \cdots, k \end{array} \right\}$$

since when μ_i is positive we can be in error only when $x_i \leqq -h$ and when μ_j is negative we can be in error only when $x_j \geqq h$ Since every event covered in (A 2) is also covered in (A 3), the probability (A 3) is $\geqq$ than the probability (A 2) The probability of avoiding error is thus $\geqq 1 - \alpha$ and the probability of error in consequence $\leqq \alpha$

REFERENCES

[1] The *Biometrika* Office, *Tables of the Incomplete Beta Function*, London, University College (1948)

[2] Duncan, D B, "A significance test for differences between ranked treatments in an analysis of variance," *Virginia Journal of Science*, 2 (new series) (1951), 171.

[3] Dunnet, C W, and Sobel, M., "A bivariate generalization of student's t-distribution with tables for certain special cases," *Biometrika*, 41 (1954), pp 153–69

[4] Feller, W, *An Introduction to Probability Theory and Its Applications*, New York, John Wiley and Sons (1950), p 75

[5] Hartley, H O, and Pearson, E S, "Table of the probability integral of the t-distribution," *Biometrika*, 37 (1950), 168–72.

[6] Kimball, A. W, "On dependent tests of significance in the analysis of variance," *Annals of Mathematical Statistics*, 22 (1951), 600–2

[7] Nair, K. R, "Distribution of the extreme deviate from the sample mean," *Biometrika*, 35 (1948), 118–44.

[8] National Bureau of Standards, "Tables of normal probability functions," *Applied Mathematical Series*, 23 (1953)

[9] Snedecor, G W., *Statistical Methods*, 4th edition, Ames Iowa, The Collegiate Press (1946), p 266 and p. 282

[10] Thompson, W R, "On a criterion for the rejection of observations and the distribution of the ratio of deviation to sampling standard deviation," ***Annals of Mathematical Statistics***, 6 (1935), 214–19

ESTIMATION OF THE PARAMETERS OF A SKEWED DISTRIBUTION BY LINEAR SYSTEMATIC STATISTICS

A. E. SARHAN
University of North Carolina

1 INTRODUCTION

IN RECENT literature, linear combinations of the sample ordered values are used to provide estimates of the parameters of certain distributions [3, 4, 5, 10]. These estimates and the general class of statistics which are derived from order statistics are termed systematic by Mosteller [6]. The efficiencies of these estimates and of some other linear estimates have been discussed for certain symmetric distributions [3, 7, 10].

It has been shown, in the case of the rectangular population [5, 10], that the two extreme values (with equal numerical weights) provide us with the best linear estimate of the mean and standard deviation. In Sections 2 and 3 below it is shown that the two extreme sample elements will have the largest weights in the best linear estimates of the mean and standard deviation of a given skewed distribution with finite range. It is also shown that for small samples the midrange and the range (which are based on the two extreme sample elements) as estimates of the mean and standard deviation for the given distribution can be used without loss of efficiency The efficiency of these estimates, and of some other linear estimates, is discussed in Sections 4 and 5.

In making the computations, the variances and covariances of the order statistics were first computed as described in Section 2. The method of least squares [1, 9] was then used to obtain the best linear estimates of the parameters of the distribution In applying this method to problems of this kind, it is convenient to employ matrices and vectors to compute the coefficients in the estimating functions and the variance of the resulting estimates The operations are outlined in Sections 6 and 7

2 THE BEST LINEAR ESTIMATES OF THE PARAMETERS OF AN ILLUSTRATIVE SKEWED DISTRIBUTION

Consider the skewed distribution,

$$(2.1)\qquad \frac{12}{\theta_2}\left(\frac{y-\theta_1}{\theta_2}+\frac{2}{3}\right)^2\left(\frac{1}{3}-\frac{y-\theta_1}{\theta_2}\right),\quad \theta_1-\frac{2\theta_2}{3}\leqq y\leqq\theta_1+\frac{\theta_2}{3},$$

where θ_1 is the true mode and θ_2 is the true range. Let

$$y_1, y_2, y_3, \cdots, y_n \tag{2.2}$$

denote a sample of size n drawn from this distribution and ordered so that

$$y_1 < y_2 < y_3 \cdots < y_n. \tag{2.3}$$

Now consider the following linear combination of these ordered sample values.

$$\theta_j^* = \sum_{i=1}^{n} \alpha_{ji} y_i, \qquad j = 1, 2. \tag{2.4}$$

where θ_j^* is an estimate of the parameter θ_j The coefficients in (2 4) can be found by the method of least squares so as to minimize the sampling variance; i e , so that θ_j^* is the best linear estimate of the jth parameter.

Let

$$x = \frac{y - \theta_1}{\theta_2} + \frac{2}{3} \cdot \tag{2.5}$$

Then the distribution of x is

$$f(x) = 12\, x^2(1 - x), \qquad 0 \leqq x \leqq 1. \tag{2.6}$$

For the rth order statistic from distribution (2 6), the moment of order k about the origin is defined to be

$$E(x_r^k) = \frac{n!}{(r-1)!(n-r)!} \int_0^1 x_r^k \left[\int_0^{x_r} f(x)dx\right]^{r-1} f(x_r)\left[\int_{x_r}^1 f(x)dx\right]^{n-r} dx_r, \tag{2.7}$$

where E stands for expected value. The second moment about the true mean is

$$V(x_r) = E(x_r^2) - [E(x_r)]^2, \tag{2.8}$$

where V stands for variance. The product moment about the origin, for the order statistics x_r and x_s where $x_r < x_s$, is defined to be

$$E(x_r x_s) = \frac{n!}{(r-1)!(s-r-1)!(n-s)!} \int_{x_s=0}^1 \int_{x_r=0}^{x_s} x_r x_s \left[\int_0^{x_r} f(x)dx\right]^{r-1} f(x_r)\left[\int_{x_r}^{x_s} f(x)dx\right]^{s-r-1} f(x_s)\left[\int_{x_s}^1 f(x)dx\right]^{n-s} dx_r dx_s \tag{2.9}$$

The product moment about the true means is

$$\text{Cov}\,(x_r x_s) = E(x_r x_s) - E(x_r)E(x_s), \tag{2.10}$$

where Cov denotes covariance

Finally, the relationships between the lower moments of x_r and y_r are given by

$$E(y_r) = \theta_1 + \theta_2\left[E(x_r) - \frac{2}{3}\right], \tag{2.11}$$

$$E[y_r - E(y_r)]^2 = \theta_2^2\{E(x_r^2) - [E(x_r)]^2\} \tag{2.12}$$

$$E[y_r y_s - E(y_r)E(y_s)] = \theta_2^2[E(x_r x_s) - E(x_r)E(x_s)], \quad r \neq s. \tag{2.13}$$

All the expected values and the variances and covariances for order statistics for samples from distribution (2.6) have been computed for n as large as 5. Table 1 gives the exact expected values for $n=2$, 3, 4,

TABLE 1

EXACT EXPECTED VALUES OF THE ORDER STATISTIC x_r IN SAMPLES FROM THE SKEWED DISTRIBUTION (2 6) EACH VALUE MUST BE DIVIDED BY THE APPROPRIATE DIVISOR GIVEN IN THE LAST COLUMN

n	x_1	x_2	x_3	x_4	x_5	Divisor
2	17	25				35
3	2127	3039	3843			5005
4	32799	46239	57087	68079		85085
5	115433	161449	197097	230153	265837	323323

5 Table 2 gives the exact variances and covariances The coefficients α_{1i} and α_{2i} for the best linear estimates of the mode, and the range for samples up to the size 5 are given in decimal form in Tables 3 and 4, respectively These estimates can be computed by substituting these coefficients in equation (2 4) The variances of θ_1^* and θ_2^* are shown in terms of θ_2^2.

3 THE BEST LINEAR ESTIMATE OF THE MEAN AND STANDARD DEVIATION

The mean of distribution (2 1) is $\theta_1 - \theta_2/15$, and its variance is $\theta_2^2/25$, so that estimates of the mean (μ) and standard deviation (σ) can be computed from the equations

TABLE 2

EXACT VARIANCES AND COVARIANCES OF THE ORDER STATISTIC x_r IN SAMPLES FROM THE SKEWED DISTRIBUTION (2.6)

n	r	s = 1	2	3	4	5
2	1	$113d_2$	$48d_2$			
	2		$85d_2$			
3	1	$621011d_3$	$319824d_3$	$159264d_3$		
	2		$514219d_3$	$258048d_3$		
	3			$396501d_3$		
4	1	$151151299d_4$	$83702736d_4$	$50774880d_4$	$27126528d_4$	
	2		$128959619d_4$	$78637632d_4$	$42148128d_4$	
	3			$109419411d_4$	$58948560d_4$	
	4				$86105899d_4$	
5	1	$28426707965d_5$	$16345757040d_5$	$10697825376d_5$	$6963908160d_5$	$3872558976d_5$
	2		$24580357885d_5$	$16151561280d_5$	$10539098592d_5$	$5870339904d_5$
	3			$21524077725d_5$	$14090885424d_5$	$7867132416d_5$
	4				$18594827405d_5$	$10422379968d_5$
	5					$14844421735d_5$

$d_2=1/3675$, $d_3=1/25050025$, $d_4=1/7239457225$, $d_5=1/1568066434935$

$$\mu^* = \theta_1^* - \theta_2^*/15, \tag{3.1}$$

$$\sigma^* = \theta_2^*/5 \tag{3.2}$$

These estimates are of the linear form

TABLE 3

COEFFICIENTS α_{ni} IN THE BEST LINEAR ESTIMATE OF THE MODE (θ_1) IN SAMPLES OF THE SKEWED DISTRIBUTIONS (2 1)

n	Order statistic y_1	y_2	y_3	y_4	y_5
2	208333333	791666667			
3	232616044	133288817	634095139		
4	226185200	097200059	113992303	562622438	
5	21689712	.08170106	.08012333	.10172209	51955640

$$\theta_1^* = \sum_{i=1}^{n} \alpha_{1i} y_i$$

$$\mu^* = \sum_{i=1}^{n} \beta_{1i} y_i, \tag{3.3}$$

$$\sigma^* = \sum_{i=1}^{n} \beta_{2i} y_i. \tag{3.4}$$

The coefficients β_{1i} and β_{2i} are given in decimal form in Tables 6 and 7, respectively. It can be shown that these are the coefficients for the best linear estimates of μ and σ.

TABLE 4

COEFFICIENTS α_{2i} IN THE BEST LINEAR ESTIMATE OF THE RANGE (θ_2) IN SAMPLES FROM THE SKEWED DISTRIBUTION (2.1)

n	Order statistic				
	y_1	y_2	y_3	y_4	y_5
2	−4 575000	4 575000			
3	−2 771360652	− 310130814	3.081491422		
4	−2 162396264	− 398146805	− 009109328	2.569652397	
5	−1 82792882	− 39570250	−.15297402	.09603220	2.2805734

$$\theta_2^* = \sum_{i=1}^{n} \alpha_{2i} y_i$$

4 EFFICIENCIES OF OTHER LINEAR ESTIMATES OF THE MEAN AND STANDARD DEVIATION

Tables 8 and 9 show the relative efficiencies of other illustrative estimates of the mean and standard deviation of distribution (2 1). In constructing these tables, the following sample statistics were first computed:

TABLE 5

VARIANCES OF THE ESTIMATES θ_1^* AND θ_2^* IN SAMPLES FROM THE SKEWED DISTRIBUTION (2 1), IN TERMS OF θ_2^2

n	θ_1^*	θ_2^*
2	.020138888	.531250000
3	.012478916	236345228
4	008943489	.145318720
5	006914030	102566451

TABLE 6

COEFFICIENTS β_{1i} IN THE BEST LINEAR ESTIMATE OF THE MEAN (μ) OF THE SKEWED DISTRIBUTION (2.1)

n	Order statistic				
	y_1	y_2	y_3	y_4	y_5
2	.5	5			
3	417373421	.153964205	428662378		
4	370344951	123743179	114599591	391312278	
5	33875904	10808123	.09032160	09531994	.36751820

$$\mu^* = \sum_{i=1}^{n} \beta_{1i} y_i$$

TABLE 7

COEFFICIENTS β_{2i} IN THE BEST LINEAR ESTIMATE OF THE STANDARD DEVIATION (σ) OF THE SKEWED DISTRIBUTION (2.1)

n	Order statistic				
	y_1	y_2	y_3	y_4	y_5
2	−.8750	8750			
3	−.554272130	− 062026163	616298284		
4	− 432479253	− 079629361	− 001821866	513930479	
5	− 36558576	− 07914050	−.03059480	01920644	4561145

$$\sigma^* = \sum_{i=1}^{n} \beta_{2i} y_i$$

TABLE 8

EFFICIENCIES OF OTHER LINEAR ESTIMATES OF THE MEAN IN SAMPLES FROM THE SKEWED DISTRIBUTION (2.1)

n	Sample mean ϕ_{11}^*	Adjusted median ϕ_{12}^*	Adjusted mid-range ϕ_{13}^*
2	100 00%	100 00%	100.00%
3	97 41	61.29	98 59
4	95 91	68 59	97 61
5	91 36	51 58	93 05

TABLE 9
EFFICIENCIES OF OTHER LINEAR ESTIMATES OF STANDARD DEVIATION FROM THE SKEWED DISTRIBUTION (2.1)

n	Adjusted range ϕ_{21}^*	Adjusted "normal estimate" ϕ_{22}^*	Gini's estimate ϕ_{23}^*
2	100.00%	100.00%	100.00%
3	99.57	99.57	99.57
4	98.84	98.56	97.64
5	97.96	97.40	95.51

(4.1) Mean $$\theta_{11}^* = \frac{1}{n}\sum_{i=1}^{n} y_i$$

(4.2) Median: $$\theta_{12}^* = \begin{cases} y_{1/2(n+1)} & \text{if } n \text{ is odd.} \\ \tfrac{1}{2}(y_{(1/2)n} + y_{(1/2)(n+1)}) & \text{if } n \text{ is even.} \end{cases}$$

(4.3) Midrange $$\theta_{13}^* = \tfrac{1}{2}(y_1 + y_n)$$

(4.4) Range: $$\theta_{21}^* = y_n - y_1$$

(4.5) Normal Estimate $$\theta_{22}^* = \sum_{i=1}^{n} \gamma_i y_i$$

(4.6) Gini's Estimate: $$\theta_{23}^* = \frac{7}{8}\left\{\frac{2}{n(n-1)}\left[2u - (n+1)v\right]\right\}$$

where the γ_i's in (4.5) are the coefficients such that θ_{22}^* is the best linear estimate of the standard deviation of a normal population [3], and u and v in (4.6) are defined by the equations

(4.7) $$u = \sum_{i=1}^{n} i y_i \qquad v = \sum_{i=1}^{n} y_i$$

Unbiased estimates of the mean and standard deviation of distribution (2.1) were then computed from the statistics

(4.8) $$\mu_1^* = \theta_{11}^*,$$

(4.9) $$\mu_2^* = \theta_{12}^* + C_n\theta_2^*,$$

(4.10) $$\mu_3^* = \theta_{13}^* - C_n'\theta_2^*,$$

(4.11) $$\sigma_1^* = \theta_{21}^*/\left[E(x_n) - E(x_1)\right],$$

(4.12) $$\sigma_2^* = K_n\theta_{22}^*,$$

(4.13) $$\sigma_3^* = \theta_{23}^*,$$

where C_n, C_n' and K_n are constants which can be calculated for each value of n to make the estimate unbiased.

5 DISCUSSION

Tables 3 and 6 show that for estimating the mode and the mean, the two extreme sample values should be assigned the greatest numerical weights, while the other values should have smaller weights Tables 4 and 7 show a similar situation for estimates of the range and standard deviation.

It is of interest to see that the least sample value (the extreme value on the side of the long tail) has a smaller coefficient than the largest sample value (the other extreme on the side of the shorter tail) This is to be expected, since extreme values from the longer tail occur more often and tend to upset the estimate It throws some light on the effect of the shape of the distribution (or its tails) on the coefficients of the best linear estimates. The nature of this effect presents an interesting but rather difficult problem.

Tables 8 and 9 show that for distribution (2 1), the efficiency of the sample mean decreases more rapidly than the efficiency of the normal estimate as the sample size increases, although they are the best linear estimates of the mean and standard deviation, respectively, in samples from a normal distribution. A similar observation applies to the mid-range and the range, although they are both based on the extreme sample values.

Table 8 shows that the efficiency of the midrange is higher than that of the sample mean, and that the efficiency of the median is low. So the midrange and the sample mean can be used as inefficient estimates for estimating the mean of the distribution (2 1), while the median is unreliable for such estimation

From Table 9, it is clear that the range has a high efficiency in small samples from this distribution The efficiency of the normal estimate is nearly the same as the range, but there is no advantage in using it, since the range is easier to compute, and is slightly more efficient Gini's estimate is not as efficient as the other estimates, but it may be useful in view of its very simple coefficients and the fact that it is unbiased regardless of sample size

From this investigation, it is evident that the calculations of the optimum coefficients are tedious even for small samples, and will be very difficult to work out for larger samples. It seems that the method that Jones [4] has recently published will be convenient to estimate the mode avoiding tedious calculations. In fact, he used the distribution

discussed in this article to illustrate his method calculating the coefficients for estimates of the mode for the case where $n=3$. The method used by Jones deals with estimates of location parameters only and will not be satisfactory for distributions that do not have a rounded mode.

However, after this good start by Jones it is hoped that more investigations may lead to a general theory or approximation for estimating the parameters of any given distribution.

6 THE EXTENDED METHOD OF LEAST SQUARES

When the expected values of observed random variables are linear functions of unknown parameters, the principle of least squares provides equations the solution of which leads to estimates of the parameters with certain optimum properties

Assume we are given n observations, $y_1, y_2, \cdots, y_n$, with expected values that are linear functions of $s<n$ unknown parameters $\theta_1, \theta_2, \cdots, \theta_s$, so that

$$(6.1) \qquad E(y_r) = \sum_{j=1}^{s} a_{rj}\theta_j, \qquad r = 1, 2, \cdots, n,$$

where $E(y_r)$ denotes the expected value of y_r In other words, writing A for the matrix of n rows and s columns with elements a_{rj}, and θ for the column vector with elements θ_j, let us assume that

$$(6.2) \qquad E(\underline{y}) = A\underline{\theta}.$$

Also, assume that

$$(6.3) \qquad V(\underline{y}) = V,$$

where $V(\underline{y})$ is the variance matrix of the vector $\underline{y}$, with elements known apart from a scalar factor. Then under these two conditions, (6 2) and (6.3), the extended principle of least squares selects the estimates $\theta_1^*, \theta_2^*, \cdots, \theta_s^*$ so as to minimize

$$(6.4) \qquad (\underline{y} - A\underline{\theta})'V^{-1}(\underline{y} - A\underline{\theta}),$$

where $(\underline{y}-A\underline{\theta})'$ is the transpose of $(\underline{y}-A\underline{\theta})$, and V^{-1} is the inverse of V with respect to $\theta_1, \theta_2, \cdots, \theta_s$ considered as independent variables.

Differentiating with respect to θ_r, equating to zero, and accumulating for all values of j, we get [9]

$$(6.5) \qquad A'V^{-1}\underline{y} = (A'V^{-1}A)\underline{\theta}^*;$$

that is,

$$\underline{\theta}^* = (A'V^{-1}A)^{-1}A'V^{-1}\underline{y}, \tag{6.6}$$

where $\underline{\theta}^*$ is the vector of the best linear estimates.

The variance matrix of the estimates of the parameters is given by

$$V(\underline{\theta}^*) = (A'V^{-1}A)^{-1}. \tag{6.7}$$

Furthermore, given the general solution (6.6), the extended principle asserts that if new parameters, $\phi_1, \phi_2, \cdots, \phi_s$, are introduced by the equation

$$\underline{\phi} = M\underline{\theta}, \tag{6.8}$$

where M is a matrix with s columns, then the least squares estimate of $\underline{\phi}$ is

$$\underline{\phi}^* = M\underline{\theta}^* = M(A'V^{-1}A)^{-1}A'V^{-1}\underline{y} \tag{6.9}$$

The matrix of variances and covariances of ϕ^* is evidently

$$M(A'V^{-1}A)^{-1}M'. \tag{6.10}$$

7 INVERTING THE VARIANCE MATRIX

The major part of the calculation is in inverting the variance matrix V, with elements v_{ij} equal to the variances and covariances of the order statistics The method used can be summarized as follows·

The matrix V is expressed as the product $T'T$ where T is an $n \times n$ upper-triangular matrix, and T' is its transpose, the elements of T being such that

$$t_{ii} = >0 \qquad i = 1, 2, \cdots, n, \tag{7.1}$$

$$t_{ij} = 0 \qquad \text{if } i > j, \tag{7.2}$$

$$v_{ij} = \sum_{r=1}^{i} t_{ri}t_{rj} \tag{7.3}$$

so that

$$\begin{bmatrix} v_{11} & v_{12} & v_{13} \cdots & v_{1n} \\ v_{12} & v_{22} & v_{23} \cdots & v_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ v_{1n} & v_{2n} & v_{3n} \cdots & v_{nn} \end{bmatrix}$$

$$= \begin{bmatrix} t_{11} & 0 & 0 & \cdots 0 \\ t_{12} & t_{22} & 0 & \cdots 0 \\ \cdot & \cdot & \cdot & \cdot \\ t_{1n} & t_{2n} & t_{3n} & \cdots t_{nn} \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} & t_{13} & \cdots t_{1n} \\ 0 & t_{22} & t_{23} & \cdots t_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots t_{nn} \end{bmatrix}.$$

The elements of T are calculated from

$$
(7.4)\quad \begin{cases}
t_{1k} = v_{1k}/t_{11} & (k = 1, 2, \cdots, n) \\
t_{2k} = (v_{2k} - t_{12}/t_{1k})/t_{22} & (k = 2, 3, \cdots, n) \\
t_{3k} = (v_{3k} - t_{13}t_{1k} - t_{23}t_{2k})/t_{33} & (k = 3, 4, \cdots, n) \\
\cdots \quad \cdots\cdots \quad \cdots & \\
t_{kk}^2 = v_{kk} - t_{1k}^2 - t_{2k}^2 - \cdots - t_{k-1,k}^2. &
\end{cases}
$$

As a check, we can evaluate both sides of the equation

$$
(7.5)\qquad \sum_{j=1}^{n} v_{jk} = t_{1k} \sum_{j=1}^{n} t_{1j} + t_{2k} \sum_{j=2}^{n} t_{2j} + \cdots + t_{kk} \sum_{j=k}^{n} t_{kj}.
$$

The inverse of T is an upper-triangular matrix T^{-1} whose diagonal elements are the reciprocals of the corresponding diagonal elements of T The reciprocal of V is obtained from

$$
(7.6)\qquad TV^{-1} = (T^{-1})'.
$$

In using (7 6) to calculate V^{-1}, there is no need to find T^{-1}. The number of different elements in T^{-1} is $\frac{1}{2}n(n+1)$ Of these, $\frac{1}{2}n(n-1)$ are zeros, and n are diagonal elements (each being the reciprocal of the corresponding diagonal element of T) The solution of $\frac{1}{2}n(n+1)$ equations provides us with values of $\frac{1}{2}n(n+1)$ elements of V^{-1} and the other elements can be obtained by symmetry.

The following is a simple example illustrating the method Suppose we wish to find the inverse of the symmetric matrix

$$
V = \begin{bmatrix} 7 & 0 & 28 & 0 \\ 0 & 28 & 0 & 196 \\ 28 & 0 & 196 & 0 \\ 0 & 196 & 0 & 1588 \end{bmatrix}
\qquad
\begin{array}{c} \sum_1 \\ 35 \\ 224 \\ 224 \\ 1784 \end{array}
$$

Using (7.4), we get

$$
T = \begin{bmatrix} \sqrt{7} & 0 & 4\sqrt{7} & 0 \\ 0 & 2\sqrt{7} & 0 & 14\sqrt{7} \\ 0 & 0 & 2\sqrt{21} & 0 \\ 0 & 0 & 0 & 6\sqrt{6} \end{bmatrix}
\qquad
\begin{array}{c} \sum_2 \\ 5\sqrt{7} \\ 16\sqrt{7} \\ 2\sqrt{21} \\ 6\sqrt{6} \end{array}
$$

where $\sum_1$ and $\sum_2$ are the sums of the corresponding rows of V and T and can be used in (7.6) as a check By (7.6) we have

$$\begin{bmatrix} \sqrt{7} & 0 & 4\sqrt{7} & 0 \\ 0 & 2\sqrt{7} & 0 & 14\sqrt{7} \\ 0 & 0 & 2\sqrt{21} & 0 \\ 0 & 0 & 0 & 6\sqrt{6} \end{bmatrix} \begin{bmatrix} v^{11} & v^{12} & v^{13} & v^{14} \\ v^{12} & v^{22} & v^{23} & v^{24} \\ v^{13} & v^{23} & v^{33} & v^{34} \\ v^{14} & v^{24} & v^{34} & v^{44} \end{bmatrix} = \begin{bmatrix} \frac{1}{t_{11}} & 0 & 0 & 0 \\ ? & \frac{1}{t_{22}} & 0 & 0 \\ ? & ? & \frac{1}{t_{33}} & 0 \\ ? & ? & ? & \frac{1}{t_{44}} \end{bmatrix}$$

where the v_{ij} are the elements of V^{-1}, and where the question mark (?) represents an element in $(T^{-1})'$ that need not be calculated for finding the inverse The last diagonal element in $(T^{-1})'$ is equal to the last row of T multiplied by the last column in V^{-1} Thus

$$\frac{1}{6\sqrt{6}} = 6\sqrt{6}v^{44};$$

that is,

$$v^{44} = \frac{1}{216}.$$

Proceeding up, we have ten equations in ten unknowns whose solution provides us with 10 elements of the inverse The other elements can be obtained by symmetry

The equations are in general given by

$$v^{ij} = \begin{cases} -\dfrac{\sum\limits_{r=i+1}^{n} t_{ir}v^{rj}}{t_{ii}} & \text{for } i \neq j \\[2ex] \dfrac{1}{t_{ii}^2} - \dfrac{\sum\limits_{r=i+1}^{n} t_{ir}v^{rj}}{t_{ii}} & \text{for } i = j. \end{cases} \tag{7.7}$$

The inverse of the matrix in the example, therefore is,

$$V^{-1} = \begin{bmatrix} 1/3 & 0 & -\frac{1}{21} & 0 \\ 0 & \frac{397}{1512} & 0 & -\frac{7}{216} \\ -\frac{1}{21} & 0 & \frac{1}{84} & 0 \\ 0 & -\frac{7}{216} & 0 & \frac{1}{216} \end{bmatrix}.$$

The method here described is a modification of the standard square root method [2] and involves less calculation (since the non-zero off-diagonal elements of T^{-1} are not required) It also leads to computed values of the elements in the inverse that are usually more exact.

The author would like to express his acknowledgment to Dr. B. G. Greenberg for his help and supervision.

REFERENCES

[1] Aitken, A C, "On least squares and linear combination of observations," *Proceedings Royal Society of Edinburgh*, 40 (1935), 42–48

[2] Dwyer, Paul S., *Linear Computations*, New York, John Wiley and Sons, 1951

[3] Godwin, H J, "On the estimation of dispersion by linear systematic statistics," *Biometrika*, 36 (1946), 92–100.

[4] Jones, H. L, "Approximating the mode from weighted sample values," *Journal of the American Statistical Association*, 48 (1953), 113–27

[5] Lloyd, H. E, "Least-squares estimation of location and scale parameters using order statistics," *Biometrika*, 39 (1952), 87–95

[6] Mosteller, F, "On some useful inefficient statistics," *Annals of Mathematical Statistics*, 17 (1946), 377–407

[7] Nair, K. R, "Efficiencies of certain linear systematic statistics for estimating dispersion from normal samples," *Biometrika*, 37 (1950), 182–83

[8] Nair, U S, "The standard error of Gini's mean difference," *Biometrika*, 23 (1936), 428–34

[9] Plackett, R L, Lecture notes on the method of least squares

[10] Sarhan, A E, "Estimation of the mean and standard deviation by order statistics" *Annals of Mathematical Statistics*, 25 (1954), 317–28

A FAMILY OF J-SHAPED FREQUENCY FUNCTIONS

CHESTER W. TOPP
Fenn College

AND

FRED C. LEONE
Case Institute of Technology

1. INTRODUCTION

THE usual procedure in obtaining theoretical distributions is to obtain a statement of the frequency function, $f(x)$ The distribution function, $F(x)$, may be expressed as a definite integral. One may adopt the opposite viewpoint and obtain a distribution function, $F(x)$, from which the associated frequency function, $f(x)$, may be obtained by differentiation. $F(x)$ will be referred to as a cumulative frequency function (c.f f.). The latter viewpoint was discussed by Burr in [1]

It is the purpose of this paper to give a family of c f f.'s which is believed to be new, which is readily handled from the calculational standpoint, which has useful values of α_3 and α_4, the standard third and fourth moments, including the range of values for failure data mentioned below, and which yields J-shaped frequency functions. A graph is presented which shows the values of α_3^2 and

$$\delta = \frac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3}$$

assumed by members of the family, and some examples are given which show that a satisfactory fit of empirical data may be obtained by the use of these functions

A collection of some sixty sets of empirical data with J-shaped histograms reveals several distributions with α_3 and α_4 values which fall under Pearson's Type I classification Of these, there is a definite group dealing with failures such as frequency of powered band tool failures, frequency of automatic calculating machine failures, and frequency of failure at time x of radar components These sets of failure data have α_3 values ranging from 57 to 1 66 and δ values ranging from $-$ 47 to $-$ 58 There appears to be a justification, therefore, for an examination of theoretical curves which yield such values

2. A FAMILY OF CUMULATIVE FREQUENCY FUNCTIONS

The family of c.f.f.'s considered here is defined as follows:

$$
(1)\quad
\begin{aligned}
F(x) &= \frac{a}{b^{2r}}(2bx - x^2)^r + (1-a)\frac{x}{b}, && 0 \leqq x \leqq b < \infty,\\
F(x) &= 0, && x < 0,\\
F(x) &= 1, && x > b,
\end{aligned}
$$

where $0<r<1$ and $0<a\leqq 1$

The corresponding family of frequency functions obtained by differentiation of (1) is given by

$$
(2)\qquad f(x) = \frac{2ar}{b^{2r}}(b-x)(2bx-x^2)^{r-1} + \frac{1-a}{b}.
$$

Two further differentiations yield the functions

$$
(3)\quad f'(x) = \frac{2ar}{b^{2r}}(2bx-x^2)^{r-2}[(b-x)^2(2r-1)-b^2],
$$

$$
(4)\quad f''(x) = \frac{4ar}{b^{2r}}(b-x)(r-1)(2bx-x^2)^{r-3}[(b-x)^2(2r-1)-3b^2].
$$

From (2) it is seen that $f(x)$ is positive for $0<x<b$, from (3) $f'(x)$ is seen to be negative for all values $0<x<b$, and from (4) $f''(x)$ is seen to be positive for all x such that $0<x<b$ Thus all frequency functions of the family (2) have graphs which are positive between 0 and b, have negative slopes between 0 and b, and for each of them the slope is always increasing between 0 and b. These curves thus may be called J-shaped From (2), however, it can be seen that the frequency curves of the family do not drop to zero at the right extreme of range except in the case $a=1$ Some curves of this type, having a positive ordinate at the right extreme of range, are called U-shaped by Pearson

Cumulative moments M_k are given by

$$
(5)\qquad M_k = \int_0^b x^k[(1-F(x)]dx,
$$

where the c f f $F(x)$ is used. In addition,

$$
(6)\quad
\begin{aligned}
\mu_1' &= M_0\\
\mu_2 &= 2M_1 - M_0^2\\
\mu_3 &= 3M_2 - 6M_1M_0 + 2M_0^3\\
\mu_4 &= 4M_3 - 12M_2M_0 + 12M_1M_0^2 - 3M_0^4,
\end{aligned}
$$

where the μ's are the ordinary moments about the mean for the frequency function $f(x)$.

Using (5) and some straight-forward integration, one may obtain for this family of functions

$$M_k = b^{k+1}\left[aN_k + \frac{1-a}{k+1} - \frac{1-a}{k+2}\right], \tag{7}$$

where the N_k are the cumulative moments for those members of the family (1) having $a=1$ and the range from zero to one, that is, for the c f f 's

$$F(x) = (2x - x^2)^r \qquad 0 \leqq x \leqq 1.$$

The N_k are calculated from (5) and are given by

$$N_0 = 1 - R$$

$$N_1 = \frac{1}{2} + \frac{1}{2r+2} - R,$$

$$N_2 = \frac{1}{3} + \frac{1}{r+1} - 2R + W,$$

$$N_3 = \frac{1}{4} + \frac{2}{r+1} - \frac{1}{2(r+2)} - 4R + 3W,$$

where

$$R = \frac{\pi}{2^{2r+2}} \frac{\Gamma(2r+2)}{\Gamma^2\left(\frac{2r+3}{2}\right)} = \frac{1}{2}\,\beta\left(\frac{1}{2},\, r+1\right)$$

and

$$W = \frac{\pi}{2^{2r+4}} \frac{\Gamma(2r+4)}{\Gamma^2\left(\frac{2r+5}{2}\right)} = \left(\frac{2r+2}{2r+3}\right)R.$$

Since M_k contains the factor b^{k+1} and no other terms containing b, it can be seen from (6) that α_3 and α_4 are independent of b However

$$\mu_1' = b\left(\frac{a+1}{2} - aR\right), \tag{8}$$

$$\mu_2 = b^2\left[\frac{1+2a}{3} + \frac{a}{r+1} - 2aR\right] - (\mu_1')^2. \tag{9}$$

The standard moments α_3 and α_4 thus depend on the two parameters a and r and may be computed from the following formulas with the aid of (7)

$$\alpha_3 = \frac{3M_2 - 6M_1M_0 + 2M_0^3}{(2M_1 - M_0^2)^{3/2}}, \tag{10}$$

$$\alpha_4 = \frac{4M_3 - 12M_2M_0 + 12M_1M_0^2 - 3M_0^4}{(2M_1 - M_0^2)^2} \tag{11}$$

Following Craig [2], values of α_3^2 and

$$\delta = \frac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3}$$

are used as coordinates in order to plot the members of the family of c.f.f.'s given by (1), thus affording a comparison with the Pearson system of frequency curves. The α_3^2 and δ values assumed by members of the family are given in Figure 1 This is a graph of α_3^2 and δ versus a and r. It covers a portion of the original Craig chart If $a=r=1$, the frequency function is triangular and results in the minimum $\delta = -40$. An upper bound for δ may be found by taking $a=1$ and letting r approach 0 This yields a δ value of approximately -81 It might be pointed out here that each point on the chart corresponds to an infinite number of members of the family (1) obtained by varying b For a given empirical distribution, after the sample α_3^2 and δ are calculated, an a and r can be selected from the contour lines in the graph A more complete comparison between α_3^2 and δ and a and r is presented in Table 1. A linear interpolation may be used between values in the table

3. FITTING EMPIRICAL DATA

The method of fitting a theoretical c f f to a given set of data has been described by Burr in [1] and by Hatke in [5] For the family of c f f 's given by (1), a brief description of the method is given here. Consider only those members of the family where $b=1$, i e consider the family

$$\begin{aligned} F(x) &= a(2x - x^2)^r + (1 - a)x, && 0 \leqq x \leqq 1, \\ F(x) &= 0, && x < 0, \\ F(x) &= 1, && x > 1. \end{aligned} \tag{12}$$

Calculate α_3 and α_4 for the given set of data and use Figure 1 or Table 1 to obtain values of a and r to use in (12). For this chosen c.f.f., calculate

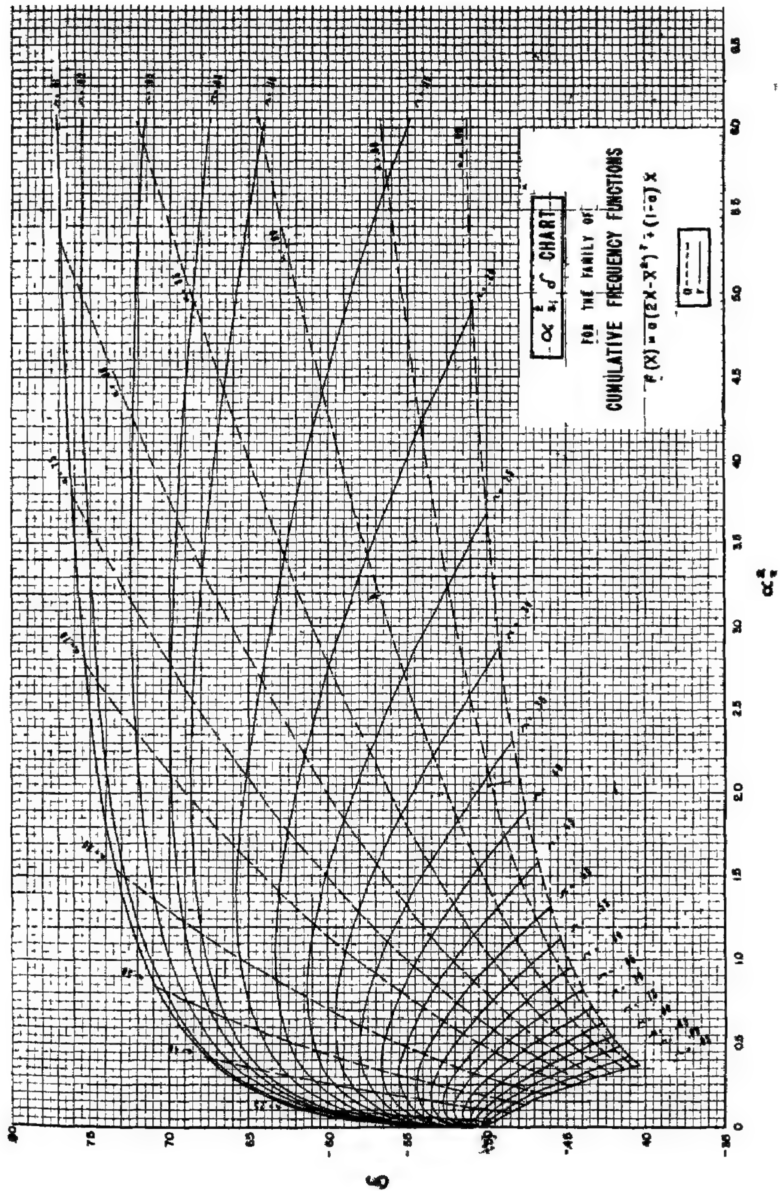

α_3^2, δ CHART
FOR THE FAMILY OF
CUMULATIVE FREQUENCY FUNCTIONS
$F(X) = a(2X - X^2)^r + (1-a)X$

TABLE 1

VALUES OF α_3^2, δ FOR a, r $F(x)=a(2x-x^2)^r+(1-a)x$

a/r	01		02		05		08		.09	
	α_3^2	δ	α_3^2	δ	α_3^2	δ	α_3^2	δ	α_3^2	δ
05	000	− 513	000	− 513	001	− 513	001	− 512	001	− 512
10	006	− 534	006	− 537	006	− 535	006	− 533	006	− 532
15	023	− 562	021	− 564	021	− 559	021	− 556	021	− 554
20	052	− 592	052	− 589	050	− 583	048	− 578	048	− 576
25	103	− 616	103	− 615	097	− 606	093	− 599	091	− 596
30	179	− 639	175	− 635	165	− 626	157	− 617	153	− 615
35	283	− 659	277	− 655	260	− 644	244	− 634	239	− 631
40	422	− 677	412	− 673	384	− 661	359	− 649	351	− 645
45	604	− 694	589	− 689	545	− 675	507	− 662	495	− 658
50	840	− 709	816	− 703	752	− 688	694	− 674	676	− 669
55	1 144	− 722	1 110	− 716	1 015	− 699	931	− 683	905	− 678
60	1 541	− 734	1 490	− 727	1 352	− 708	1 231	− 691	1 194	− 685
.65	2 066	− 745	1 992	− 738	1 788	− 716	1 614	− 696	1 562	− 690
70	2 780	− 754	2 667	− 745	2 365	− 721	2 111	− 699	2 035	− 692
75	3 790	− 762	3 612	− 752	3 147	− 724	2 767	− 699	2 655	− 691
80	5 302	− 769	5 004	− 756	4 253	− 723	3 661	− 695	3 492	− 687
85	7 780	− 773	7 233	− 758	5 910	− 717	4 935	− 685	4 663	− 675
90	12 530	− 774	11 310	− 752	8 624	− 700	6 846	− 663	6 382	− 652
95	25 110	− 760	21 010	− 725	13 720	− 657	9 911	− 617	9 026	− 607
1 00	124 000	− 649	61 230	− 547	23 600	− 540	14 220	− 534	12 480	− 532

a/r	10		15		20		25		30		35	
	α_3^2	δ	α_3^2	δ	α_3^2	δ	α_3^2	δ	α_3^2	δ	α_3^2	δ
05	001	− 512	001	− 512	001	− 511	001	− 510	001	− 510	002	− 509
10	006	− 531	007	− 529	007	− 526	007	− 524	007	− 522	007	− 520
15	021	− 553	020	− 548	020	− 543	020	− 538	020	− 535	019	−.531
20	047	− 574	045	− 566	043	− 559	041	− 553	040	− 547	038	− 542
25	090	− 594	084	− 584	079	− 574	074	− 566	070	− 558	066	− 551
30	151	− 612	139	− 599	128	− 588	119	− 577	111	− 568	104	− 559
35	234	− 628	213	− 613	195	− 599	178	− 588	164	− 576	152	− 566
40	343	− 642	309	− 625	280	− 610	254	− 596	232	− 583	212	− 572
45	483	− 654	430	− 635	386	− 618	348	− 603	315	− 589	286	− 576
50	659	− 664	582	− 644	518	− 625	463	− 608	416	− 593	375	− 579
55	881	− 673	771	− 650	[illegible]	− 630	602	− 611	.536	− 595	480	− 580
60	1 159	− 680	1 004	− 655	876	− 632	770	− 613	680	− 595	604	− 579
65	1 512	− 684	1 292	− 657	1 115	− 632	970	− 612	850	− 593	749	− 575
70	1 963	− 686	1 652	− 656	1 407	− 630	1 210	− 608	1 050	− 588	916	−.570
75	2 550	− 685	2 106	− 652	1 766	− 625	1 498	− 601	1 284	− 580	1 109	− 562
80	3 334	− 679	2 684	− 643	2 206	− 614	1 842	− 590	1 556	− 569	1 329	− 550
85	4 415	− 666	3 434	− 628	2 750	− 599	2 249	− 574	1 870	− 554	1 575	− 536
90	5 967	− 642	4 418	− 604	3 416	− 575	2 723	− 552	2 221	− 534	1 841	− 518
95	8 264	[illegible]	5 678	− 564	4 194	− 542	3 242	− 525	2 583	− 510	2 105	− 498
1 00	11.090	− 530	6 948	− 519	4 894	− 509	3 674	− 500	2 873	− 491	2 309	− 484

TABLE 1—(*Continued*)

a/τ	40		45		50		55		60		65	
	α_3^2	δ	α_3^2	δ	α_3^2	δ	α_3^2	δ	α_3^2	δ	α_3^2	δ
.05	002	− 508	002	− 508	002	− 507	002	− 507	002	− 506	002	− 505
10	007	− 518	007	− 516	007	− 515	007	− 512	007	− 511	007	− 510
15	019	− 528	018	− 524	018	− 521	017	− 518	016	− 516	016	− 513
20	037	− 537	035	− 532	034	− 528	032	− 523	030	− 520	029	− 516
25	062	− 545	059	− 539	056	− 533	053	− 528	050	− 523	047	− 519
30	097	− 552	091	− 544	085	− 538	080	− 531	074	− 526	070	− 520
35	141	− 557	130	− 549	121	− 541	113	−.534	105	− 527	098	− 521
40	195	− 562	180	− 552	166	− 543	153	− 535	141	− 527	131	− 520
45	261	− 565	238	− 554	218	− 544	200	− 535	184	− 527	170	− 519
50	340	− 566	308	− 555	280	− 544	256	− 534	234	− 525	214	− 516
55	432	− 566	389	− 554	352	− 542	319	− 531	290	− 522	264	− 513
60	539	− 564	483	− 551	434	− 539	391	− 528	353	− 517	319	− 508
65	663	− 560	589	− 546	526	− 534	471	− 522	423	− 512	380	− 502
70	804	− 554	710	− 540	629	− 527	560	− 515	499	− 504	447	− 495
75	965	− 546	844	− 531	743	− 519	654	− 506	581	− 496	517	− 486
80	1 144	− 535	992	− 521	866	− 508	759	− 496	668	− 485	590	− 476
85	1 340	− 521	1 151	− 508	994	− 495	865	− 485	756	− 474	663	− 466
.90	1 547	− 505	1 313	− 493	1 124	− 481	970	− 472	841	− 463	733	− 454
95	1 744	− 489	1 464	− 478	1 242	− 468	1 062	− 461	914	− 452	792	− 444
1 00	1 892	− 475	1 575	− 467	1 326	− 460	1 126	− 453	964	− 446	830	− 440

a/τ	70		75		80		85		90		95	
	α_3^2	δ	α_3^2	δ	α_3^2	δ	α_3^2	δ	α_3^2	δ	α_3^2	δ
05	002	− 504	002	− 504	002	− 503	001	− 502	001	− 501	001	− 501
10	006	− 508	006	− 506	006	− 504	006	− 502	005	− 502	005	− 500
15	015	− 511	014	− 508	014	− 506	013	− 504	012	− 502	012	− 500
20	028	− 513	026	− 510	025	− 507	023	− 504	022	− 501	020	− 498
25	044	− 514	042	− 510	039	− 507	037	− 502	034	− 499	032	− 496
.30	065	− 515	061	− 510	057	− 506	053	− 501	050	− 497	.046	− 493
35	091	− 515	084	− 509	079	− 504	073	− 499	068	− 494	062	− 490
40	121	− 514	112	− 507	104	− 502	096	− 496	088	− 491	082	− 486
45	156	− 512	144	− 505	132	− 498	122	− 492	112	− 486	103	− 481
50	196	− 509	179	− 501	163	− 494	151	− 488	138	− 482	126	− 475
55	240	− 504	219	− 496	200	− 489	182	− 482	166	− 476	152	− 470
60	289	− 499	262	− 491	238	− 483	216	− 476	197	− 469	179	−.463
65	342	− 492	310	− 484	280	− 476	252	− 469	229	− 462	207	− 456
70	400	− 485	359	− 476	323	− 469	291	− 461	262	− 453	236	− 447
75	461	− 476	412	− 468	368	− 460	330	− 452	295	− 445	265	− 439
80	522	− 467	464	− 458	413	− 451	368	− 443	328	− 437	293	− 430
85	.584	− 456	516	− 449	456	− 441	405	− 435	359	− 428	319	− 422
.90	641	− 446	562	− 439	495	− 432	437	− 425	386	− 419	341	− 414
95	689	− 438	601	− 431	526	− 424	462	− 420	406	− 414	357	− 408
1 00	.718	− 433	624	− 427	544	− 421	475	− 416	416	− 410	364	− 404

μ_1' and σ by use of (8) and (9). Change the scale and origin of the values of the variable X of the given data to those x's corresponding to (12) by means of

$$\frac{x-\mu_1'}{\sigma}=\frac{X-\overline{X}}{s}, \tag{13}$$

where $\overline{X}$ and s are the mean and standard deviation of the given data. The x values to use in (12) are thus seen to be

$$x=\frac{\sigma}{s}X+\left(\mu_1'-\frac{\sigma}{s}\overline{X}\right). \tag{14}$$

The X class limits are substituted into (14) and corresponding values of (12) are calculated and differenced to give the probabilities for the given ranges of X These probabilities may be multiplied by the total frequency to yield theoretical frequencies

An example of the fitting process is given here using Davis' data [3] on the lifetime in hours of transmitter tubes used in aircraft radar sets [1]

For this distribution $\overline{X}=149\ 480$, $s=139\ 539$, $\alpha_3{}^2=1\ 587$ and $\delta=-\ 505$ are obtained empirically The graduation was done by using $r=40$ and $a=91$ in (12). These values were read from Figure 1 which was entered with the values of $\alpha_3{}^2$ and δ Thus, the c f f. used for graduation was

$$F(x)=.91(2x-x^2)^{\ 40}+.09x. \tag{15}$$

By using equations (8) and (9) and $b=1$ we get

$$R=\frac{\pi}{2^{2\ 8}}\frac{\Gamma(2\ 8)}{\Gamma^2(1\ 9)}=.817578,$$

$$\mu_1'=\frac{a+1}{2}-aR=.211005,$$

$$\sigma=\left[\left(\frac{1+2a}{3}+\frac{a}{r+1}-2aR\right)-(\mu_1')^2\right]^{1/2}=.239776.$$

Substitution in equation (14) gives us

$$x=.0017183X-.045844$$

---

[1] Davis presented thirty distributions Approximately fifteen distributions were tested. Half of these fell into the range of this J-shape, and better than half of these proved to have a good fit by this distribution.

TABLE 2

HOURS BEFORE FAILURE OF TRANSMITTER TUBES

Lifetime in hours X	Class limits in X	Class limits in x	$F(x)$	Graduated frequency by $F(x)$	Graduated frequency by exponential [3]	Observed frequency
25				112 0	94 7	100
	50	.04006	33247			
75				63 9	68.1	68
	100	12598	.52217			
125				38 5	50 2	48
	150	21191	63632			
175				28 3	34 0	31
	200	29783	72028			
250				40 2	43.8	42
	300	.46968	.83967			
350				26 7	22 2	21
	400	.64153	.91903			
500				27 3	24 0	27
	600	1 00000	1 00000			
Total				336.9	337 0	337

into which one may substitute the X class limits 50, 100, 150, etc. The values of x obtained are substituted into (15) to calculate the $F(x)$ values which when multiplied by 337 and differenced give the theoretical frequencies found in column 5 of Table 2

If the χ^2 comparison test is used, the fit obtained by the use of $F(x)$ is good One obtains $\chi^2=5\ 499$ whereas $P(\chi^2>4.878)=.30$ and $P(\chi^2>5.989)=20$ for 4 degrees of freedom The corresponding figures for the exponential graduation as given in [3] are $\chi^2=1\ 13$ and $P=95$ for 5 degrees of freedom

4. SOME EXAMPLES

In this section a few examples are given to show the satisfactory fit which may be obtained with $F(x)$ The first example is taken from [3] and concerns the lifetime in hours for 100, V600 indicator tubes used in aircraft radar sets.[2] The c.f f. used for this graduation was

[2] The value, $X=1000$, in Table 3 differs from that indicated in Davis' original paper in which the highest interval was left open. The actual data were obtained from the author and this last interval is closed.

TABLE 3

HOURS BEFORE FAILURE OF V600 INDICATOR TUBE

Lifetime in hours X	Observed frequency	Graduated frequency by $F(x)$	Graduated frequency by exponential [3]
50	29	33 7	28.8
150	22	18 3	20.0
250	12	10 9	14 8
350	10	9 8	10 2
500	10	10 0	12 8
700	9	9.9	6 5
1000	8	7 3	6 9
Total	100	99 9	100 0

$$F(x) = 83(2x - x^2).36 + .17x$$

and the fit may be measured by $\chi^2 = 1\ 649$, giving $P \cong 80$ for 4 degrees of freedom The exponential graduation gave $\chi^2 = 2\ 48$, $P \cong 78$ for 5 degrees of freedom

The last example is taken from the data given by Elderton in [4] These data relate to six months' experience of maturities among endowment assurances and are graduated by Elderton with a Pearson Type VIII curve

TABLE 4

MATURITIES AMONG ENDOWMENT ASSURANCES

X	Observed frequency	Graduated frequency by $F(x)$	Graduated frequency by Type VIII [4]
1	469	468 1	437
2	186	193.4	222
3	166	153 4	165
4	134	137 3	136
5	122	127 3	120
6	112	109 5	109
Total	1189	1189 0	1189

The c.f.f. used for the graduation was

$$F(x) = .46(2x - x^2)^{.20} + .54x$$

and yields $\chi^2 = 1.677$, whereas $P(\chi^2 > 1.424) = .70$ and $P(\chi^2 > 2.366) = .50$ for 3 degrees of freedom.

REFERENCES

[1] Burr, Irving W., "Cumulative frequency functions," *Annals of Mathematical Statistics*, 13 (1942), 215–32

[2] Craig, C C , "A new exposition and chart for the Pearson system of frequency curves," *Annals of Mathematical Statistics*, 7 (1936), 16–28.

[3] Davis, D. J., "An analysis of some failure data," *Journal of the American Statistical Association*, 47 (1952), 113–50.

[4] Elderton, W P , *Frequency Curves and Correlation*, fourth edition, Washington, D. C , Harren Press (1953).

[5] Hatke, M. A., "A certain cumulative probability function," *Annals of Mathematical Statistics*, 20 (1949), 461–63.

BASIC PROBLEMS, TECHNIQUES, AND THEORY OF ISOPLETH MAPPING*

CALVIN F. SCHMID AND EARLE H. MACCANNELL
Office of Population Research, University of Washington

THIS paper is concerned with the basic problems, techniques and theory of isopleth mapping. The term "isopleth," Gr. isoplethes, equal in quantity or number, (from isos-equal and plethos-quantity, number) as used in this discussion, designates one of two types of isoline maps in which the lines (isopleths) connect equal rates or ratios for specific areas In the other type of isoline map, which is commonly referred to as an isometric map, lines (isometers) are drawn through points of equal value or intensity In the "isopleth" map, the values are rates or ratios computed for areal units, such as census tracts, townships, precincts, or counties whereas in the "isometric" map the values are samples of absolute measurement taken at different points on a map A population density map with lines showing equal densities is an example of the "isopleth" map, and a topographic map with lines connecting a series of points of equal elevation ("isohypses," or more commonly, contour lines) is an example of the "isometric" map [1]

Although there have been several important contributions to the literature on problems of isopleth mapping, all have failed in varying degrees (1) to state certain of the problems definitively and with reference to basic statistical principles, (2) to offer adequate solutions to the stated problems, and (3) to formulate clearly the relationship between statistical theory and isopleth mapping Obviously, the present paper makes no pretense of presenting the final word on these three desiderata, but it does attempt to make more clear and explicit certain of the basic problems, solutions, and statistical theory of isopleth mapping.[2]

---

* Acknowledgment is made to Lloyd Kirry and Gloria Austin for their assistance in drafting the charts. The authors also are indebted to their colleagues, Z William Birnbaum and Douglas G Chapman of the Department of Mathematics, John Sherman of the Department of Geography, and Stuart C Dodd and George A. Lundberg of the Department of Sociology for a critical reading of the manuscript

[1] It will be found that the terminology used in connection with both "isopleth" and "isometric" maps as designated above, is inconsistent, if not confusing For example, instead of the generic term, "isoline" many cartographers prefer "isogram" (equal-line), "isopleth" (equal-measure), "isorithm" (equal-number), and a few, "isometric" and "isontic "

[2] Perhaps the most relevant and significant discussions in this connection are Fr Uhorczak, "Metoda Izarytmiczna W Mapach " *Polski Przeglad Kartograficzny* Tom IV (Grudzien 1929), 95–124, and J. Ross Mackay, "Some problems and techniques in isopleth mapping," *Economic Geography*, XXVII (January, 1951), 1–9.

SUMMARY OF PROBLEMS OF ISOPLETH MAPPING

The major problems of isopleth mapping may be summarized as follows:

1. What influence does the size of base areas have on isopleths? Is there an "optimum-sized" base area for isopleth maps?
2. What is the most logical "control point" and how is it located?
3. What is the logic and technique of determining class-intervals for isopleths?
4. What interpolation technique is most appropriate for isopleth maps? What factors influence the reliability of interpolation?
5. Are there any principles derivable from statistical theory that might be useful in the formulation of a sound rationale of isopleth mapping as well as in providing a basis for clarifying and perhaps solving some of the problems discussed in this paper?

IMPLICATIONS OF SIZE OF BASE AREA

Since the data for isopleth maps are based on defined areas, their size and shape exert a pronounced influence on the reliability, comparability, significance, and general appearance of the isopleth map If the base areas are relatively large, meaningful variations are masked and the isopleths are extremely general. On the other hand, if the areas are relatively small, chance and possible meaningless variations in the data will be recorded as myriads of tiny "islands" or "peaks" on the isopleth map [3]

Sometimes the cartographer has a limited choice of areas since the data may be classified in two, three, or even four ways For example, data for a city may be classified on the basis of blocks, enumeration districts, census tracts or community areas, and for a state, into enumeration districts or precincts, census divisions or townships, and counties The size of the territorial units in relation to the entire area to

[3] These "islands" or "peaks" may be meaningless in two ways First, if the variation within the blocks, or other small areas, is not significantly less than the variation between these areas and some larger areas, such as enumeration districts, the variations which are recorded on the basis of the small areas are not statistically significant Second, although the variations indicated by small areas are statistically significant, it may be unnecessary or even confusing for many purposes to use the smallest significant base areas

Some statistical technique, such as analysis of variance, will be found useful in determining the minimum significant size of base areas, even though it may not be desirable to use such small areas in every case The base areas of blocks, enumeration districts and census tracts used in Figures 1, 2 and 3 were analyzed by analysis of variance techniques The variation within about 80 per cent of the blocks was found to be significantly less than the variation between the blocks within their respective enumeration districts. The variation within all but one of the enumeration districts was found to be significantly less than the variation between enumeration districts within their respective census tracts The variation within the census tracts was found to be significantly less than the variation for the whole segment included on the maps. The criterion of the significance as used here was defined at the five per cent level.

be mapped, the nature of the data, and particularly the purpose at hand should be the major determinants in selecting a particular type of area For example, the authors have constructed more than 15 isopleth maps for the City of Chicago (approximately 207 square miles of land area) based on approximately 1,000 census tracts in 1940. If block data or enumeration data had been used, there would have been many thousands of little islands which would have resulted in excessive and unnecessary detail On the other hand, community areas with their relatively large size and heterogeneity would have concealed important and significant patterns.

A comparison of three isopleth maps for the northern and eastern segment of Seattle based on (1) city blocks, (2) enumeration districts, and (3) census tracts respectively are shown in Figures 1, 2, and 3. The map comprises approximately 20 square miles of land area out of a total land area of 70.8 square miles for the City of Seattle in 1950 In the area covered by these maps there are approximately 2,400 blocks, 290 enumeration districts, and 36 census tracts After carefully examining Figures 1, 2, and 3 it is difficult, if not impossible, to say which is "best," since any judgment of this kind is usually related to the purpose for which the map is designed. However, for certain purposes it can be said that the isopleths based on block data provide too much detail, while those based on census tract data are overgeneralized. In this connection, the size of the base unit is always relative to the over-all area included on the map

It must not be overlooked that the problem of size and shape of the areal units has another implication with respect to isopleth mapping When the base areas vary widely in size a loss in measurability of values occurs in the interpolation process If one control point is based on a relatively large area, and the other on a small area, the interpolation of values between them is subject to errors of interpretation which cannot be controlled statistically If it is impossible or impracticable to use base areas of approximately the same size and shape, extreme caution should be followed in interpreting the resulting isopleths

LOCATION OF CONTROL POINT

In constructing an isopleth map a point is used to represent each areal unit This point which is called the control point of the area, must be accurately located according to some specific assumption for each areal unit for which a rate or ratio has been derived If the unit is symmetrical in shape and the distribution of the phenomenon is relatively uniform, the geographical center naturally would be the control point.

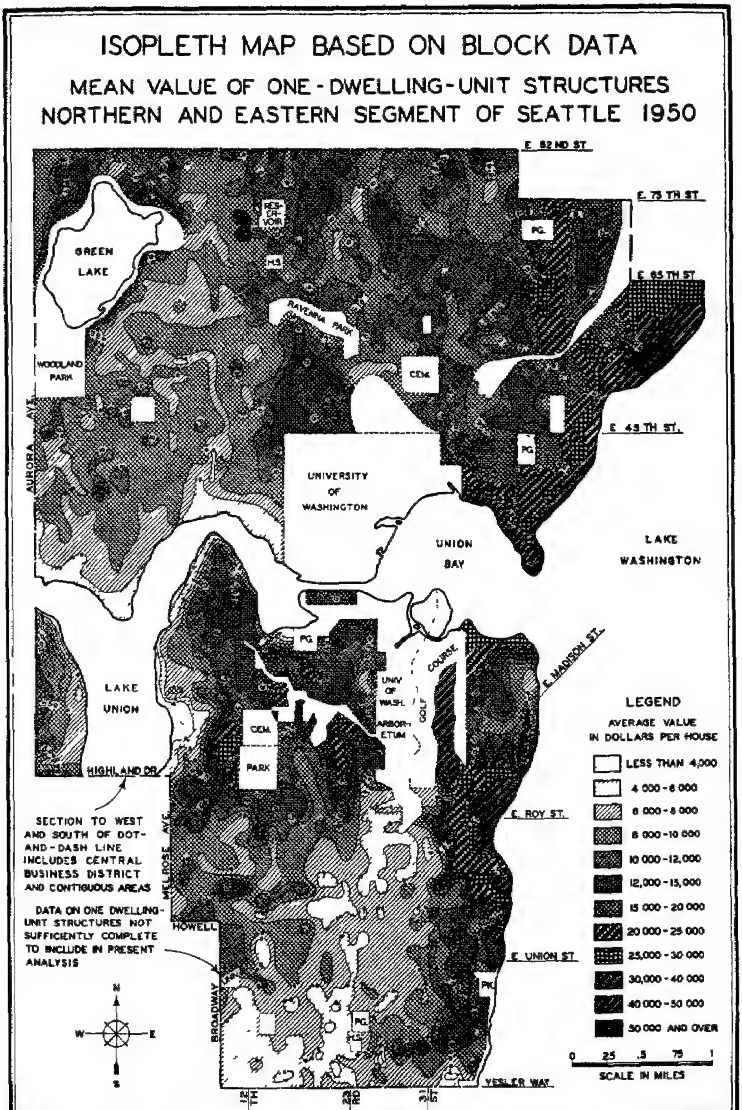

ISOPLETH MAP BASED ON BLOCK DATA
MEAN VALUE OF ONE-DWELLING-UNIT STRUCTURES
NORTHERN AND EASTERN SEGMENT OF SEATTLE 1950
E. 82 ND ST
E. 75 TH ST
E. 65 TH ST
E. 45 TH ST.
RES-ER-VOIR
H.S.
P.G.
GREEN LAKE
RAVENNA PARK
WOODLAND PARK
CEM.
AURORA AVE.
UNIVERSITY OF WASHINGTON
UNION BAY
LAKE WASHINGTON
E. MADISON ST.
LAKE UNION
GOLF COURSE
UNIV. OF WASH. ARBOR-ETUM
PARK
HIGHLAND DR.
E. ROY ST.
MELROSE AVE.
HOWELL
E. UNION ST.
BROADWAY
PK.
YESLER WAY
12 TH
23 RD
31 ST
SECTION TO WEST AND SOUTH OF DOT-AND-DASH LINE INCLUDES CENTRAL BUSINESS DISTRICT AND CONTIGUOUS AREAS
DATA ON ONE DWELLING-UNIT STRUCTURES NOT SUFFICIENTLY COMPLETE TO INCLUDE IN PRESENT ANALYSIS
N
W
E
S
LEGEND
AVERAGE VALUE IN DOLLARS PER HOUSE
LESS THAN 4,000
4 000 - 6 000
6 000 - 8 000
8 000 - 10 000
10 000 - 12,000
12,000 - 15,000
15 000 - 20 000
20 000 - 25 000
25,000 - 30 000
30,000 - 40 000
40 000 - 50 000
50 000 AND OVER
0 .25 .5 .75 1
SCALE IN MILES

FIGURE 1

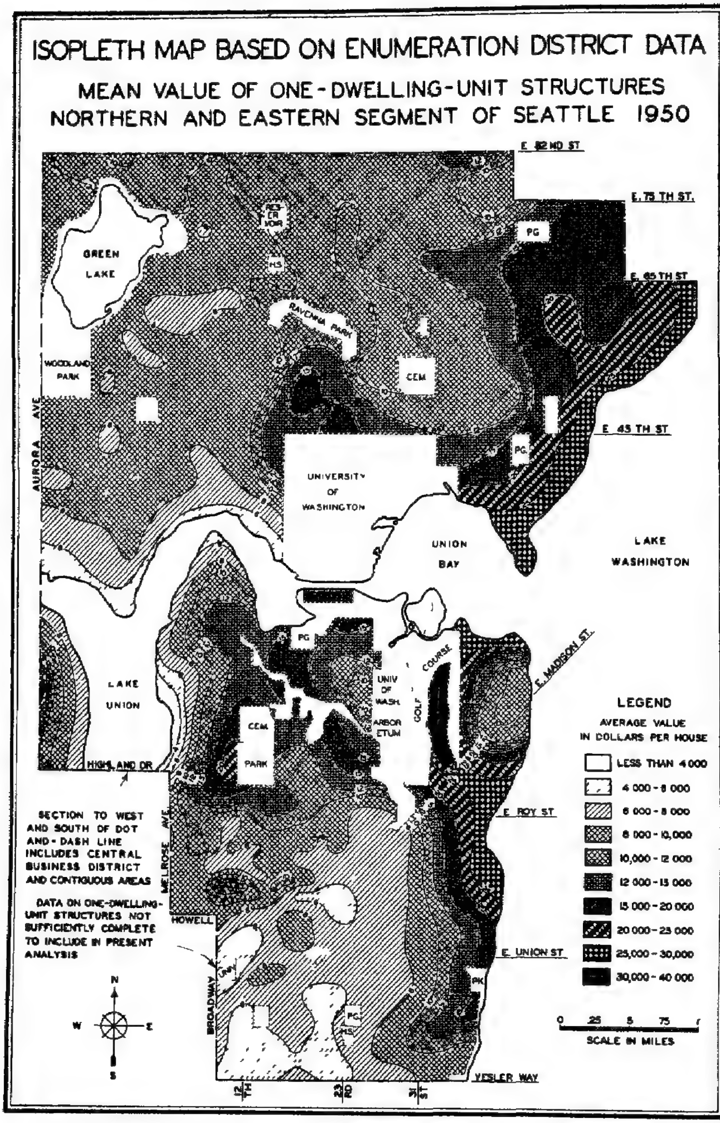

ISOPLETH MAP BASED ON ENUMERATION DISTRICT DATA
MEAN VALUE OF ONE-DWELLING-UNIT STRUCTURES
NORTHERN AND EASTERN SEGMENT OF SEATTLE 1950
E. 75 TH ST.
E. 65 TH ST
E. 45 TH ST
GREEN LAKE
RAVENNA PARK
WOODLAND PARK
CEM
PG
HS
AURORA AVE
UNIVERSITY OF WASHINGTON
UNION BAY
LAKE WASHINGTON
LAKE UNION
HIGHLAND DR
UNIV OF WASH. ARBOR ETUM
GOLF COURSE
E. MADISON ST.
PARK
E. ROY ST
E. UNION ST
MELROSE AVE
HOWELL
BROADWAY
YESLER WAY
PK
SECTION TO WEST AND SOUTH OF DOT AND-DASH LINE INCLUDES CENTRAL BUSINESS DISTRICT AND CONTIGUOUS AREAS
DATA ON ONE-DWELLING-UNIT STRUCTURES NOT SUFFICIENTLY COMPLETE TO INCLUDE IN PRESENT ANALYSIS
N
W
E
S
LEGEND
AVERAGE VALUE IN DOLLARS PER HOUSE
LESS THAN 4 000
4 000 - 6 000
6 000 - 8 000
8 000 - 10,000
10,000 - 12 000
12 000 - 15 000
15 000 - 20 000
20 000 - 25 000
25,000 - 30,000
30,000 - 40 000
SCALE IN MILES

FIGURE 2

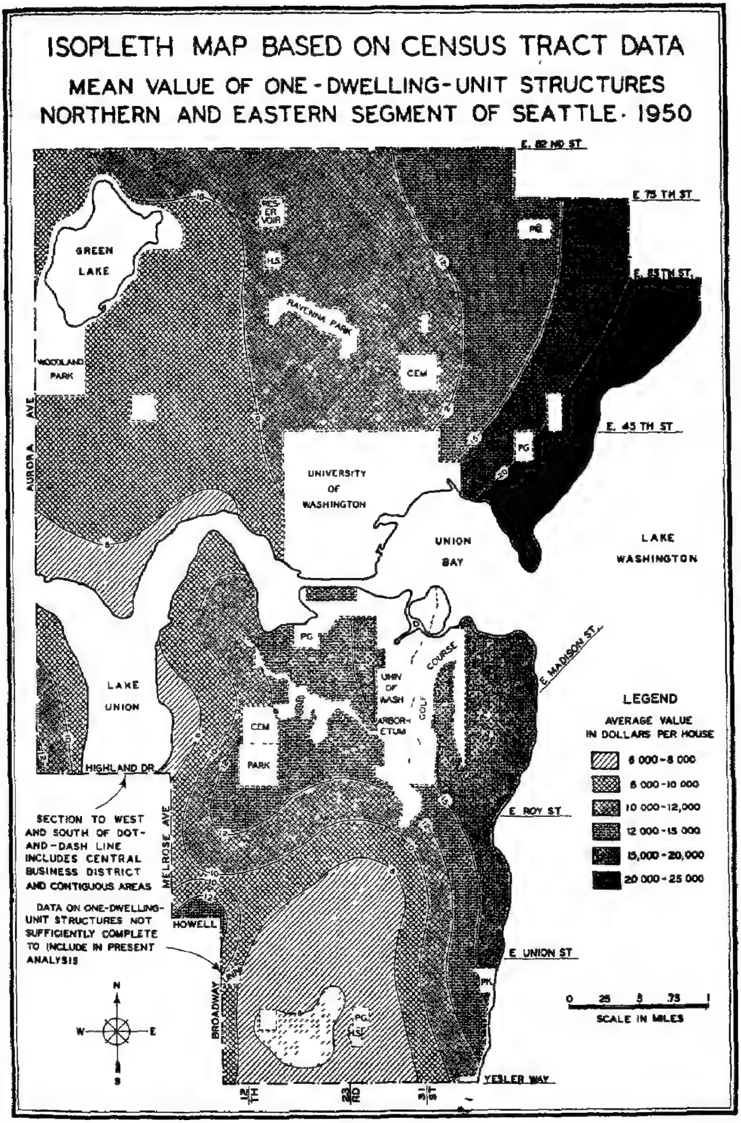

ISOPLETH MAP BASED ON CENSUS TRACT DATA
MEAN VALUE OF ONE-DWELLING-UNIT STRUCTURES
NORTHERN AND EASTERN SEGMENT OF SEATTLE · 1950
E. 82 ND ST
E 75 TH ST
E. 65TH ST.
E. 45 TH ST
GREEN LAKE
WOODLAND PARK
AURORA AVE
RAVENNA PARK
CEM
UNIVERSITY OF WASHINGTON
UNION BAY
LAKE WASHINGTON
LAKE UNION
HIGHLAND DR.
E. MADISON ST.
GOLF COURSE
UNIV OF WASH ARBORETUM
PARK
E ROY ST
E UNION ST
YESLER WAY
MELROSE AVE
HOWELL
BROADWAY
SECTION TO WEST AND SOUTH OF DOT-AND-DASH LINE INCLUDES CENTRAL BUSINESS DISTRICT AND CONTIGUOUS AREAS
DATA ON ONE-DWELLING-UNIT STRUCTURES NOT SUFFICIENTLY COMPLETE TO INCLUDE IN PRESENT ANALYSIS
LEGEND
AVERAGE VALUE IN DOLLARS PER HOUSE
6 000-8 000
8 000-10 000
10 000-12,000
12 000-15 000
15,000-20,000
20 000-25 000
0 .25 .5 .75 1
SCALE IN MILES
N
S
E
W

FIGURE 3

In actuality, census tracts, precincts, counties, and other areas for which statistics are compiled are seldom symmetrical, and the patterns of distribution of most economic, social, and other phenomena are uneven The problem is to determine the most typical or representative point in each area This point in any type of area, with even or uneven distribution of some characteristic, is the center of gravity or pivot point where the distribution would balance if it were supported by a rigid and weightless plane [4]

Occasionally, in unusually shaped areas and uneven distributions the control point may be located outside the boundaries of area.

DETERMINING SIZE OF INTERVALS

The third consideration is the choice of intervals for isopleths The isopleth interval may be based on either a geometric or arithmetic progression or some division or multiple of a number such as, 5 or 10, resulting, for example, in line values, 0, 5, 10, 25, 50, 100, 500, or 1,000 The size of the interval should be adapted to each map, particularly in relation to its purpose as well as to the type of distribution, reliability, and other characteristics of the data If the intervals are too large, the results may be an overgeneralized and somewhat meaningless map. On the other hand, if the isopleths are plotted in accordance with small class intervals or when a map has widely separated control points, an unwarranted impression of precision is conveyed The value of each isoline is indicated on the map with an appropriate number and/or by a hatching scheme, the significance of which is included in a supplementary legend [5]

INTERPOLATION PROBLEMS AND TECHNIQUES

One of the most frequently discussed problems in isopleth mapping pertains to interpolation procedure Although more or less complicated mathematical interpolation techniques have been proposed,[6] simple

---

[4] In mathematical terminology this point is known as the centroid of the base area "Centroid" may be defined as the point at which an area must be supported in order to balance perfectly if the area itself is a weightless plane and the elements described by the centroid are distributed over it In physics this point is known as the "first moment" or "center of gravity" and may be found mathematically by locating the mean distance of all the elements from any pair of arbitrarily chosen perpendicular axes Cf, Mackay, *op cit* and Uhorczak, *op cit*

[5] J W Alexander and G A Zahorchak, "Population-density maps of the United States techniques and patterns," *Geographical Review*, 33 (1943), 458–60, J Ross Mackay *op cit*

[6] Uhorczak devotes most of his paper (*op cit*) to the problem of interpolation from a non-linear point of view He applies logarithmic and geometric techniques in order to maintain uniformity of distribution within the areal subdivisions of a map Uhorczak assumes that the maintenance of the uniformity of distribution within the base areas is the major consideration in constructing a statistically measurable isopleth map The present writers disagree with this assumption Rather than uniformity of distribution it can be demonstrated empirically that values of most characteristics tend to approach the average of adjacent areas at the boundaries between the areas

linear interpolation seems to be most appropriate as judged by reliability of results, to say nothing of the additional time required for the more complicated techniques.[7]

Although linear interpolation in itself is not difficult, sometimes conditions arise which create serious problems The most common problem of this kind occurs whenever more than three centers are so located that interpolation axes cannot be drawn from each center to all the other

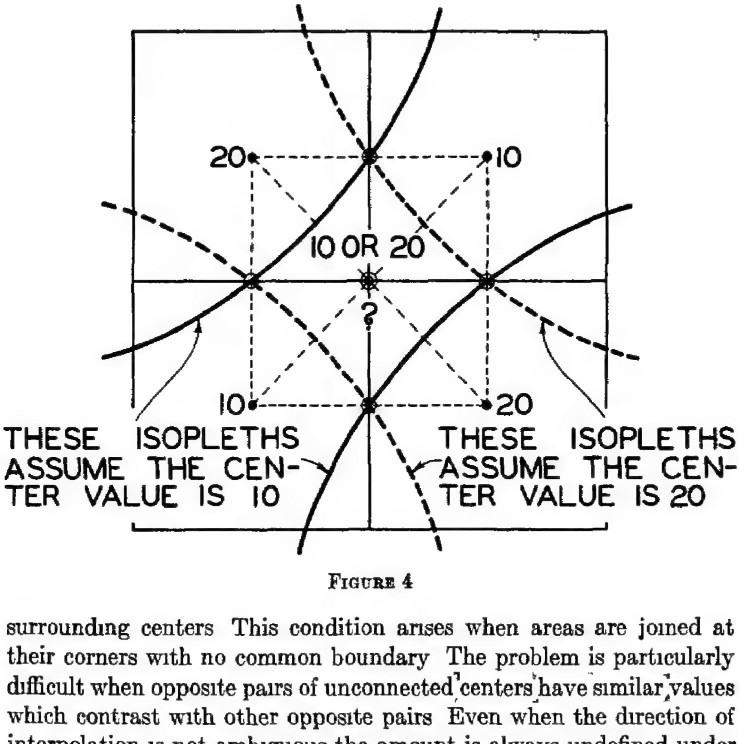


FIGURE 4

surrounding centers This condition arises when areas are joined at their corners with no common boundary The problem is particularly difficult when opposite pairs of unconnected centers have similar values which contrast with other opposite pairs Even when the direction of interpolation is not ambiguous the amount is always undefined under this condition (Figure 4).

In order to avoid this ambiguity it is necessary that the interpolation axes form a network of triangles over the map This triangular pattern will occur if the corners of adjacent areas are not joined In other words all adjacent areas must have a common boundary and not merely touch at the corners. This arrangement of fields is illustrated by Figures 5 and 9.

[7] For a discussion of linear interpolation as applied to isopleth mapping, see Calvin F Schmid, *Handbook of Graphic Presentation*, Ronald Press, New York, 1954, 212-19.

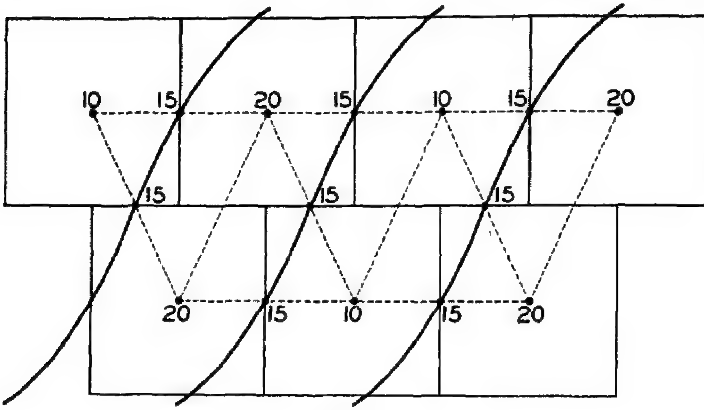


FIGURE 5

This principle of interpolation follows from the fact that the number of points of equal interpolation value around a polygon is limited by the number of sides. There is always an even number of equal-value interpolation points around a polygon and there can be no more such points than the number of sides of the polygon The interpolation axes always form polygons, and when an isoline enters a polygon by crossing a point having a given value on one of its sides, there should be one and only one other point—on one other side—by which the line could leave. Triangles fulfill this requirement Four or more sides may allow three, five, or more possible points at which the isoline may pass out of the polygon (Figure 6). Fortunately, for practical purposes, whenever adjacent fields have segments of common boundaries the three sided, or non-ambiguous, triangular network of interpolation axes results Whenever practicable, it is advisable then that a "staggered" grid of fields be developed over which to locate centers

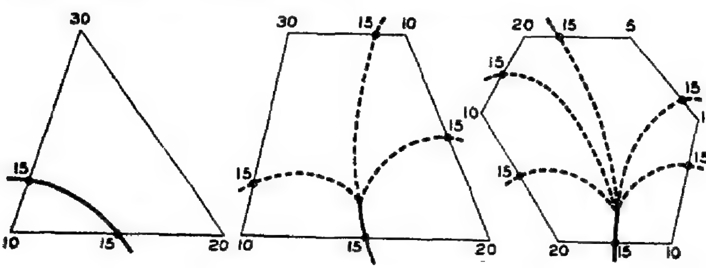


FIGURE 6

STATISTICAL THEORY AND ISOPLETH MAPPING

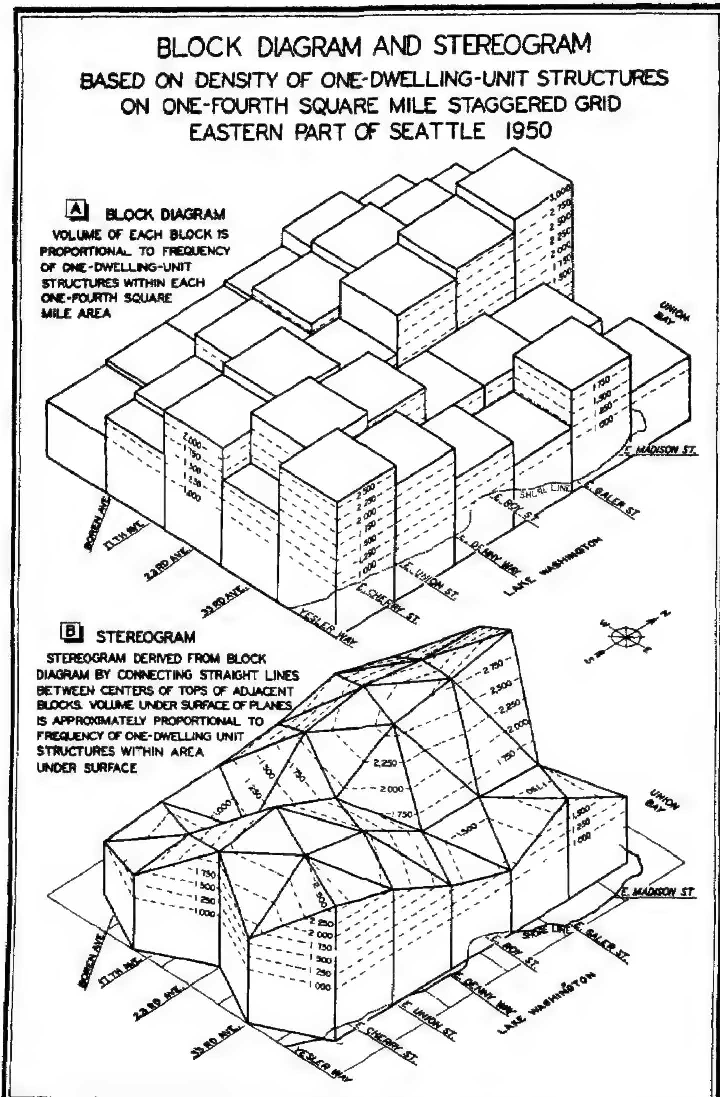
In attempting to clarify as well as to formulate solutions to the problems outlined in the foregoing paragraphs, it would seem particularly appropriate to examine the basic statistical assumptions and implications of isopleth mapping techniques.

Broadly speaking, the statistical map, including the cross-hatched as well as the isopleth map, is related to frequency distributions. In this general comparison, the cross-hatched area unit map is analogous to the histogram and the isopleth map, to the frequency polygon and smoothed frequency curve.

The map, however, is a three-dimensional representation Since the map must be presented on a flat surface—excluding the use of three-dimensional models—and the two areal dimensions are given a surface orientation, the third or characteristic dimension is portrayed as a vertical projection on the surface. The pattern map of areal subdivisions can be considered then, to be the vertical projection of the surfaces of the tops of prismatic volumes over an area If presented in isometric projection rather than vertical projection, the pattern map would appear as a block diagram (Figure 7 A).

The same assumptions and restrictions that apply to the histogram also apply to the pattern subdivision map In the histogram it must be assumed that the frequency within a given class interval is uniform over the entire interval This assumption is reflected by the horizontality of the line across the interval of a histogram at the height above the base that corresponds to this assumed uniform value In the same way, it must be assumed that the distribution of a characteristic on a pattern map is uniform over the entire area of a subdivision in order to justify the use of a uniform hatching or color pattern

The frequency represented by a histogram is proportional to the area under the curve (curve is used here in the mathematical sense, referring to the vertical and horizontal "steps" of the histogram). If the class intervals are not equal, the height of the "steps" cannot be proportional to the frequencies of the several class intervals since in a histogram the height of a column divided by the length of its base is proportional to its frequency It follows that the value of a subarea on a pattern map is proportional to the volume of a prismatic column under the subarea. Furthermore, unless all the subareas are of equal size the heights of the columns above the base plane cannot be proportional to the values of the subareas. In actual practice of course, the subareas are not usually equal, and accordingly the height is proportional to the value divided by the base area.



BLOCK DIAGRAM AND STEREOGRAM
BASED ON DENSITY OF ONE-DWELLING-UNIT STRUCTURES
ON ONE-FOURTH SQUARE MILE STAGGERED GRID
EASTERN PART OF SEATTLE 1950
A BLOCK DIAGRAM
VOLUME OF EACH BLOCK IS PROPORTIONAL TO FREQUENCY OF ONE-DWELLING-UNIT STRUCTURES WITHIN EACH ONE-FOURTH SQUARE MILE AREA
UNION BAY
E. MADISON ST.
SHORE LINE
E. GALER ST.
E. BOY ST.
E. DENNY WAY
LAKE WASHINGTON
E. UNION ST.
E. CHERRY ST.
YESLER WAY
33 RD AVE.
23 RD AVE.
12 TH AVE.
BOREN AVE.
B STEREOGRAM
STEREOGRAM DERIVED FROM BLOCK DIAGRAM BY CONNECTING STRAIGHT LINES BETWEEN CENTERS OF TOPS OF ADJACENT BLOCKS. VOLUME UNDER SURFACE OF PLANES IS APPROXIMATELY PROPORTIONAL TO FREQUENCY OF ONE-DWELLING UNIT STRUCTURES WITHIN AREA UNDER SURFACE
UNION BAY
E. MADISON ST.
SHORE LINE
E. GALER ST.
E. BOY ST.
E. DENNY WAY
LAKE WASHINGTON
E. UNION ST.
E. CHERRY ST.
YESLER WAY
33 RD AVE.
23 RD AVE.
12 TH AVE.
BOREN AVE.

FIGURE 7

Although these assumptions are useful for certain purposes, there is another approach which approximates more closely most empirical conditions. This approach is based on the logic and assumptions of the frequency polygon rather than the histogram. The area under a frequency polygon which has been constructed by connecting the midpoints of the tops of histogram rectangles is approximately equal to the area under the histogram for any class interval with the exception of a modal or antimodal interval.[8]

This relationship is also applicable to rate maps. It will be observed that the volume under the stereogram surface in Figure 7 B, for example, made up of planes connecting centroidal points on the top surfaces of the prismatic block diagram, is approximately equal to the volume of the block diagram, and with the same exceptions of modal or antimodal intervals It may be noted here that for each antimode there must be two modes (assuming a recession in values at the extremes) and that the biases tend to be compensating up to the limit of a single mode or antimode A multi-modal curve or surface is no more biased than a unimodal distribution It follows that little distortion of the representation of values—that is, statistical measurability—accrues if a block diagram is transformed into a stereogram by connecting the centroids of adjacent subareas with straight lines forming sloping planes. In addition, under the assumption that empirical agreement with data is improved by interpolation between centers, the surface of the stereogram is a better representation of the data than is the block diagram.

Although the stereogram devised on the basis of rates has no "real" or observable counterpart—the "elevations" being hypothetical constructs designed to depict the varying rates and not the result of measurements at particular points—the hypothetical elevations on the surface of the stereogram may be represented on a vertical projection by a series of curves representing equal "elevations," or actually, equal rates. The stereogram surface, if it is not smoothed, is represented by a configuration of intersecting planes (Figure 7 B) Equal elevation is represented by a plane parallel to the base of the stereogram, and such a plane will cut the stereogram surface in a series of straight line segments which form polygons in the vertical projection (Figure 8 A).

[8] The mode is a maximum. That is, any point, any area of uniform height, any histogram interval, or any other graphically described position which is surrounded on all sides by positions of lesser value is called a mode An antimode is a minimum in the same sense An antimode is surrounded on all sides by greater values Distributions may have any number of modes and antimodes. In two dimensional presentation, if the data recede at both extremes there is always one more mode than antimodes. If the data increase at the extremes there is one more antimode than modes, and if the data increase at one extreme and recede at the other there are the same number of modes as antimodes In three dimensional presentation the same relationship between modes and antimodes usually obtains.

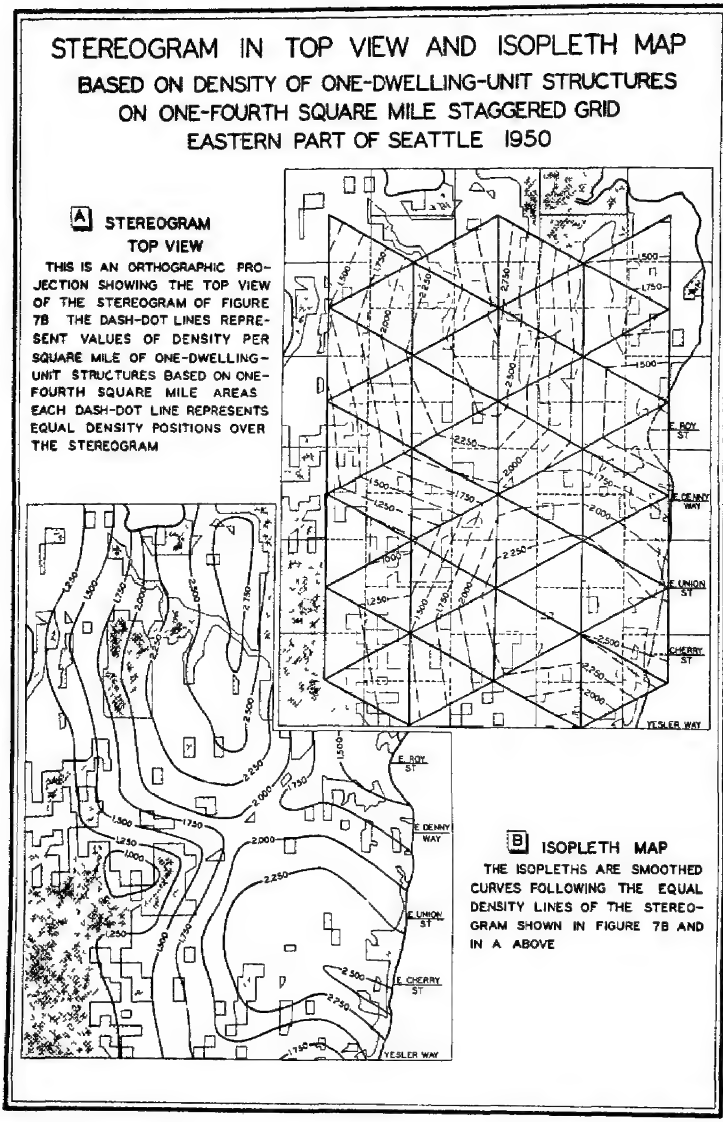

STEREOGRAM IN TOP VIEW AND ISOPLETH MAP
BASED ON DENSITY OF ONE-DWELLING-UNIT STRUCTURES
ON ONE-FOURTH SQUARE MILE STAGGERED GRID
EASTERN PART OF SEATTLE 1950
A STEREOGRAM
TOP VIEW
THIS IS AN ORTHOGRAPHIC PROJECTION SHOWING THE TOP VIEW OF THE STEREOGRAM OF FIGURE 7B THE DASH-DOT LINES REPRESENT VALUES OF DENSITY PER SQUARE MILE OF ONE-DWELLING-UNIT STRUCTURES BASED ON ONE-FOURTH SQUARE MILE AREAS EACH DASH-DOT LINE REPRESENTS EQUAL DENSITY POSITIONS OVER THE STEREOGRAM
E. ROY ST
E. DENNY WAY
E. UNION ST
CHERRY ST
YESLER WAY
B ISOPLETH MAP
THE ISOPLETHS ARE SMOOTHED CURVES FOLLOWING THE EQUAL DENSITY LINES OF THE STEREOGRAM SHOWN IN FIGURE 7B AND IN A ABOVE
E. ROY ST
E. DENNY WAY
E. UNION ST
E. CHERRY ST
YESLER WAY

FIGURE 8

These polygons can be smoothed by precise techniques, but errors involved in the various assumptions leading to their construction are sufficiently great to make precise smoothing comparable to computations beyond the limit of significant digits in elementary mathematics.[9] (Figure 8 B) In order to obviate a false appearance of accuracy and at the same time reduce chance fluctuations in the data, freehand smoothing can be used. As a test of the reliability, four draftsmen independently smoothed freehand a given polygon and the results were compared. The four curves drawn by the draftsmen were superimposed to show general agreement or disagreement and the areas measured with a planimeter. In the light of this experiment, it was found that the variations were relatively small involving a maximum difference in area of less than one per cent.

The vertical dimension on an isopleth map is assumed to vary linearly from isoline to isoline and on the basis of this assumption, approximate values for any desired area of the map can be computed By using a planimeter the desired areas can be measured, and average heights determined The value for a specified area is the area multiplied by its average height in terms of the isolines that cover it.

Frequently, the characteristic portrayed on an isopleth map is expressed as a ratio of two non-areal units, such as dollars per person (income), persons per household, or some other combination of social or economic factors When isoline maps are used to delimit these higher-ordered characteristics, the immediate relationship to area is obscured. The fact that a map is used implies some inherent association between the characteristics and their distribution over an area The association can be expressed mathematically and hence manipulated statistically if each characteristic is related separately to the area Since reciprocal values for any characteristic do not alter the isopleth placement, any line that represents a value per unit area can be said to represent also a unit area per unit value That is, a line which represents 100 persons per square mile can also be said to represent one one-hundredths square miles per person Multiplication of isopleth line values is accomplished by superimposing the corresponding isolines on the same map. If one of the sets of isolines is then expressed as a reciprocal relationship with area, the superimposed lines will cancel out the areal unit, leaving the ratio as an expression between characteristics To illustrate, a map may be prepared with isolines showing single unit dwelling structure per

---

[9] The contour polygons need not be drawn in actual practice The vertices can be connected with a smooth curve without first completing the edges of the polygons, which are not used in the final drawing anyway.

square mile. A corresponding map might show number of dollars valuation of one-dwelling-unit structures per square mile Both of these ratios are compatible with the usual concept of areal based, three-dimensional isopleth maps If one of them, let us say, dwelling units per square mile, is designated as a reciprocal—square miles per dwelling unit—the intersection of an isopleth describing one-hundred thousand dollars per square mile with an isopleth indicating 085 square miles per dwelling unit represents 100,000 times 085 or $8,500 per one-dwelling-unit structure at that point It should be pointed out that different sized areal bases for calculating the location of isolines will give different patterns for the same data. In order that the approximation be as statistically measurable as possible, it is important that the areas used to compute the isolines for both characteristics be the same size and shape The isopleth map showing mean value of one-dwelling-unit structures in the Eastern Part of Seattle in 1950 shown in Figure 9 was constructed in this way.

TECHNIQUE FOR CONSTRUCTING A STATISTICALLY VALID ISOPLETH MAP

The value for any center or control point has significance only in relation to the size and shape of the area which it represents If points along an interpolation axis are to be determined between centers representing different sized and shaped base areas, or if a center is determined by the intersection of two isolines which represent different sized or shaped base areas, the resulting points cannot be interpreted, since the base with which they must be associated cannot be designated When unequal or non-congruent base areas are used in constructing isopleths, the resulting map cannot be interpreted precisely nor considered statistically measurable Therefore, a map of this kind can be interpreted only if it is assumed to be similar to a map constructed over congruent base areas

In order to construct an "ideal" isopleth map, the characteristic or characteristics under consideration must be exactly located on a map of suitable size and projection One method by which a complete distribution of the characteristics can be exactly located is by means of a dot map Also, in order to provide an infinite number of centers or control points, it must be possible to apply without restriction as to location, a constant base area to the map. The control points thus derived represent centroids and values with reference to the distribution of the characteristics within the base area for any given points If a very large number of such centers is located, equal values will appear as a

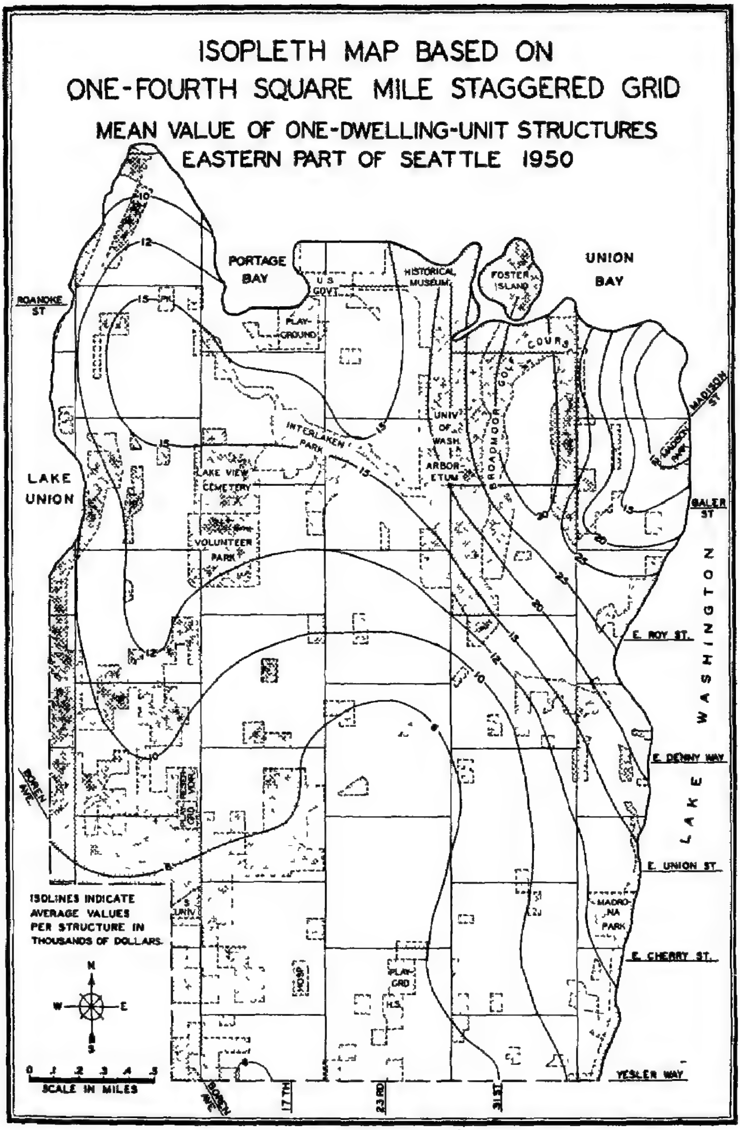

ISOPLETH MAP BASED ON
ONE-FOURTH SQUARE MILE STAGGERED GRID
MEAN VALUE OF ONE-DWELLING-UNIT STRUCTURES
EASTERN PART OF SEATTLE 1950
PORTAGE BAY
U.S. GOVT.
PLAY-GROUND
HISTORICAL MUSEUM
FOSTER ISLAND
UNION BAY
ROANOKE ST
GOLF COURSE
MADISON ST
INTERLAKEN PARK
UNIV. OF WASH. ARBORETUM
BROADMOOR
LAKE UNION
LAKE VIEW CEMETERY
VOLUNTEER PARK
GALER ST
LAKE WASHINGTON
E. ROY ST.
E. DENNY WAY
E. UNION ST.
MADRONA PARK
E. CHERRY ST.
YESLER WAY
BOREN AVE.
ISOLINES INDICATE AVERAGE VALUES PER STRUCTURE IN THOUSANDS OF DOLLARS.
N
W
E
S
0 .1 .2 .3 .4 .5
SCALE IN MILES
PLAY-GRD
H.S.
HOSP.
17TH
23 RD
31 ST
10
12
15
20
25
8

FIGURE 9

continuous series of points, or a curve, on the map. Such curves are the "true" isopleths of the characteristic on the projection used with respect to the base area from which they are computed This "ideal" isopleth map can be approached by a technique which shall be referred to as the "floating grid technique," which is applied to a dot map. This technique will yield an isopleth map which is as empirically accurate and statistically measurable as is the dot map over which it is constructed. Since the dots are located to represent the total distribution, no assumption as to the uniformity of the distribution is necessary Furthermore, since intermediate points can be determined readily by slight shifts of the "floating grid," no ambiguity of points will result, nor are any assumptions needed with respect to interpolation In practice, of course, an infinite number of centers is not located, and interpolation is used, but a sufficiently large number of centers is located so that extreme generality and excessive subjective determination is eliminated, and questionable interpolations are analyzed by precise methods.

To apply the "floating-grid technique," a suitable base map is prepared on which the characteristic under consideration is located as accurately as possible with small, uniform dots In the case of higher-order isopleth maps in which more than one characteristic is considered —such as persons per house or divorces per unit population—each of the characteristics must be located separately on the map, or separate base maps must be prepared for each characteristic A suitable base area is chosen with respect to the data under consideration. For some data, areas as small as a single city block might be suitable, whereas for other data such small areas might result in superfluous and irrelevant detail. For example, with respect to population density, a vacant lot of the size of a city block or less, for most purposes might be disregarded or considered to be a chance fluctuation However, several such vacant lots within a small district, or large tracts of vacant property in any area, are usually quite significant in this respect. If this is the case, base areas for population density isopleths should not be as small as a city block, but neither should they be so large that adjacent, heavily populated areas will obscure a series of vacant blocks or large vacant tracts A base area ranging from four blocks square to a quarter of a square mile might be acceptable for an urban population density isopleth map. For states or larger regions, however, base areas of ten to one-hundred square miles might be appropriate

From the point of view of interpretation of the isopleth map, a regular geometric form is desirable for the base area. The orientation of the base should not affect the determination of the center or control point.

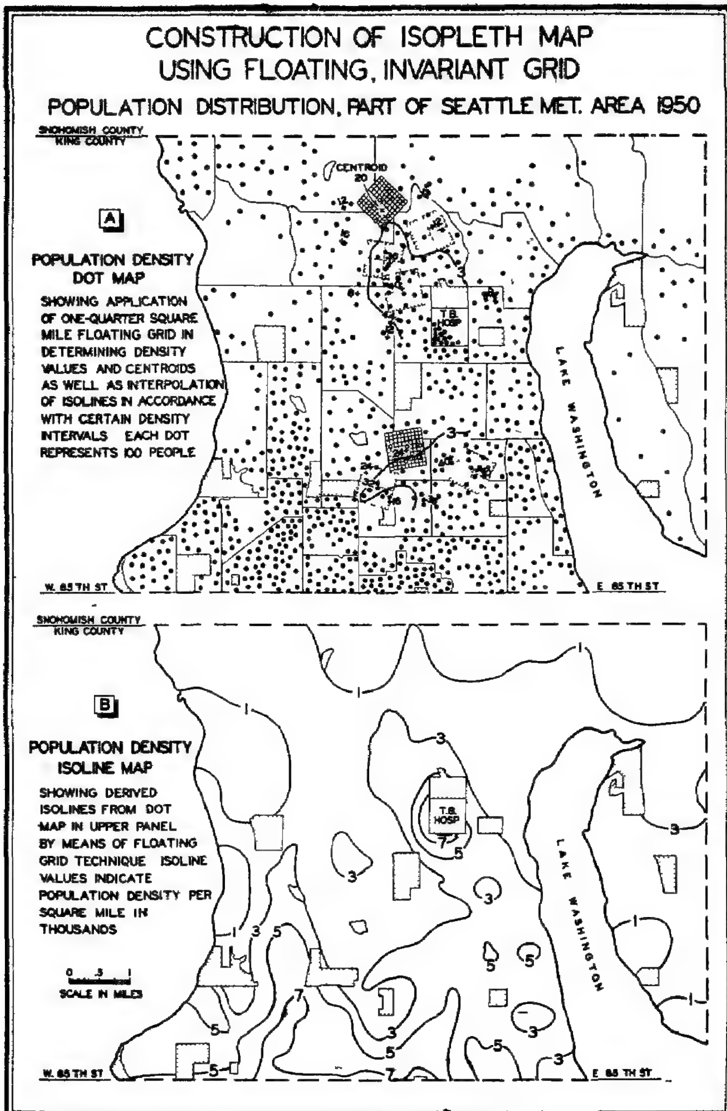

CONSTRUCTION OF ISOPLETH MAP
USING FLOATING, INVARIANT GRID
POPULATION DISTRIBUTION, PART OF SEATTLE MET. AREA 1950
SNOHOMISH COUNTY
KING COUNTY
A
POPULATION DENSITY DOT MAP
SHOWING APPLICATION OF ONE-QUARTER SQUARE MILE FLOATING GRID IN DETERMINING DENSITY VALUES AND CENTROIDS AS WELL AS INTERPOLATION OF ISOLINES IN ACCORDANCE WITH CERTAIN DENSITY INTERVALS. EACH DOT REPRESENTS 100 PEOPLE
CENTROID
T.B. HOSP
LAKE WASHINGTON
W. 85 TH ST
E. 85 TH ST
SNOHOMISH COUNTY
KING COUNTY
B
POPULATION DENSITY ISOLINE MAP
SHOWING DERIVED ISOLINES FROM DOT MAP IN UPPER PANEL BY MEANS OF FLOATING GRID TECHNIQUE. ISOLINE VALUES INDICATE POPULATION DENSITY PER SQUARE MILE IN THOUSANDS
0 .5 1
SCALE IN MILES
T.B. HOSP
LAKE WASHINGTON
W. 85 TH ST
E. 85 TH ST

FIGURE 10

In this respect any regular geometric shape—circle, triangle, pentagon, hexagon, etc.—is equally acceptable, but due to special and simple means of determining the centroid of mass within a square, the square base is recommended for practical reasons. A convenient, practical floating grid is made by drawing on some transparent material, such as transparent acetate, a square of the size determined as discussed above, and scaled to the proportions of the dot map over which it is to be applied. The square is then divided into quarters with lines parallel to its sides and each quarter marked with a fine grid of from five to ten

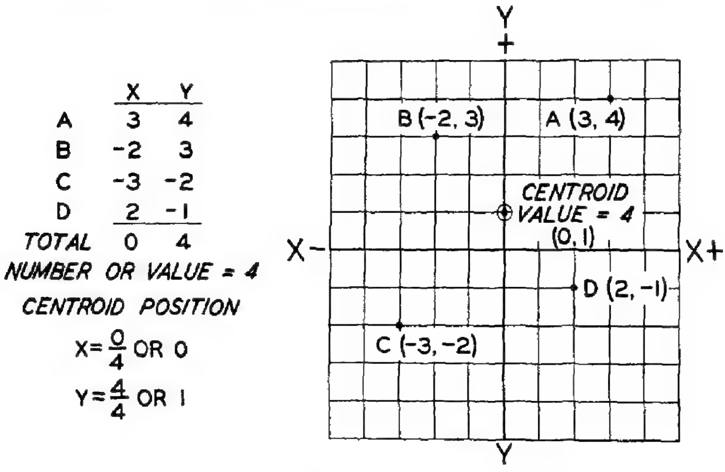


FIGURE 11

squares in each direction (Figure 10 A). The grid may be used in the same way that a rectilinear coordinate system is used to determine centroids. Each dot which falls within the grid is given an X and Y value according to its location on the coordinate system, and the mean, or average, X, Y point is the centroid of that particular base area and has a value equal to the number of dots included within the grid.

For example, in Figure 11, four dots appear within the grid. The dot marked "A" is three units up, or positive from the center and four units to the right, or positive from the center. "B" is two units to the left, or negative from the center and three units up, or positive from the center, etc. The sum of the left and right, or X direction distances is zero, and the sum of the up and down, or Y direction distances is plus 4. Since there are four dots, the mean or average distance left and right is 0/4 or zero, and the mean distance up and down is 4/4 or plus one.

The centroid, then, is located zero units left and right and one positive or up unit from the center and is given a value of four Shifting the grid to a position which would include only three of these points, if no new points were "picked up" in the area by the shift, would give a new centroid position with a value of three In this manner as many centroids or centers can be located and evaluated as are necessary to generate the desired number of isopleth lines over the map

No centers can be located on a map within a band approximately one half the width of the base grid at the limits of the map area. This band will extend along all natural boundaries as well as along vacant areas and areas adjacent to the last recorded data on the dot map. In the case of arbitrarily defined limits, this difficulty can be solved readily by including enough additional territory on the dot map so that it will be possible to complete the isopleths to the desired limits In order to extend the isolines up to lake shores or to other natural boundaries, it is necessary to proceed according to some arbitrary assumption concerning the data being depicted. First, if it is assumed that the density drops to zero rather quickly at the natural boundary, no isoline can be allowed to terminate at that boundary. For example, all lines approaching a lake shore would tend to run parallel to and near the shore line. A second, and perhaps more acceptable assumption would be to extend the values to the natural boundary and then terminate them abruptly This assumption requires the extrapolation of values across the band in which centers cannot be located Vacant property can be handled in exactly the same way as lakes, bays and rivers This procedure was used in the construction of the isoline maps in this article

No attempt should be made to locate an isoline which has a value based on less than three dots. For example, if each dot represents 100 persons, no isoline should be drawn to represent less than 300 persons per unit area where the unit area is the floating-grid size If the grid is one-quarter square miles and the dots represent 100 persons, the minimum isoline value should be 1,200 persons per square mile This principle has been violated in the case of the isolines designated *1* in Figure 10 B, and therefore the isopleth indicating 1,000 persons per square mile on this map is not as reliable as the other isopleths. When a large number of centers and their values have been located, any intermediate points and values can be determined by linear interpolation with relatively small loss in accuracy, especially if all unusual patterns of dots on the base map have been covered carefully by measured values Connecting equal-valued centers by smooth, continuous curves can be accomplished readily by a draftsman without resort to mathematical or mechanical smoothing techniques and with high reliability

THE 1955 ECONOMIC REPORT OF THE PRESIDENT*

BERYL WAYNE SPRINKEL
Harris Trust and Savings Bank, Chicago

I. THE CONTENT OF THE REPORT

All who are concerned with the role of economics in formulating public policy, as well as the manifestations of such policy in the ebb and flow of economic conditions, should welcome the 1955 *Economic Report of the President* as a clear statement of principles guiding the present Administration in dealing with economic activity Although other factors have undoubtedly affected some policy decisions, it is clear that economic analysis has played an important role. The *Report* is oriented toward an analysis of recent business developments and the outlook for 1955 with policy prescriptions for facilitating the growth of private enterprise and increasing the stability of our economy It is well written and, consequently, can be understood by the lay reader as well as by the technical economist

The *Report* consists of three chapters and four appendices Chapter 1, "The Expansive Power of the American Economy," includes a discussion of Federal Government obligations under the Employment Act, factors contributing to the achievement of growth potentials, and Government action taken during 1954 to build a stronger economy. Chapter 2, "A Year of Economic Transition," includes a detailed analysis of the nature of the business contraction in 1954, reasons for the mildness of the contraction, Federal Government policies and their effects, and a listing of lessons from experience and guides to the future.

Chapter 3, "Program for Sustained Economic Progress," is devoted to a discussion of Government policies designed to promote long-term growth The emphasis on growth rather than stabilization is due to the belief that the economy is now "undergoing a cumulative expansion of some strength" which is expected to lead to a "high and satisfactory level of employment and production within the current year" (1955, p. 48) This chapter deals with policies designed to promote the spirit of enterprise, encourage foreign trade and investment, improve social security measures, improve public assets such as public works and natural resources, and increase the stability of a growing economy.

* A review article on *The Economic Report of the President*, U S. Government Printing Office, January 1955. Pp. x, 203 $0 75

The appendices include a summary of economic recommendations of the President, a factual documentation of the analytical account of economic developments in 1954 given in Chapter 2, a report on the activities of the Council during 1954, and a set of statistical tables relating to income, employment and production.

II EVALUATION

A. *Comparative Analysis*

In order to appreciate fully the significant changes that have occurred in the President's *Economic Report* in recent years, it is useful to compare the current report with the last one written by the previous Administration in January, 1953. It should be recognized that the current Administration has benefited from the experience of the earlier work so that it would be disappointing if some improvement had not been made Although opinion on the extent of the betterments may differ, depending partially on the economic philosophy of the reader, it is the opinion of the reviewer that important improvements have been accomplished.

Perhaps the most significant change is in the economic philosophy underlying the two *Reports* Although the current *Report* does not recommend a return to laissez-faire, it places considerably more reliance on private initiative for achieving the goal of maximizing the satisfaction of human wants, and less emphasis on Government intervention The following basic economic tenets found in the current *Report* outline the Economic philosophy of this Administration. "First, competitive markets, rather than Governmental directives, are as a rule the most efficient instruments for organizing production and consumption Second, a free economy has great capacity to generate jobs and incomes if a feeling of confidence in the economic future is widely shared by investors, workers, businessmen, farmers, and consumers Third, the Federal Government creates an atmosphere favorable to economic activity when it encourages private initiative, curbs monopolistic tendencies, whether of business or labor, avoids encroachment on the private sector of the economy, and carries out as much of its own work as is practicable through private enterprise Fourth, the Federal Government generates confidence when it restrains tendencies toward recession or inflation, and does this by relying largely on indirect means of influencing private behavior rather than by direct controls over people, industries, and markets Fifth, the Federal Government contributes to economic growth when it takes its part, at the side of the States, in promoting scientific research and in providing public facili-

ties, such as highways, hospitals, harbors, and educational institutions, on which the expansion of the private economy heavily rests Sixth, the Federal Government strengthens the foundations of the economy when it widens opportunity for its less fortunate citizens and, working in cooperation with the States and localities, helps individuals to cope with the hazards of unemployment, illness, old age, and blighted neighborhoods" (1955, p 2) Although the preceding Administration would probably concur in the above principles, the economic policies pursued by the two groups indicate a greater willingness on the part of the present Administration to implement this philosophy with concrete action As indicated by points four and five, this *Report* places as much if not more emphasis on non-defense spending on public facilities and social security measures as did the previous one

Both *Reports* placed considerable emphasis upon the responsibility placed on the Government by the Full Employment Act for stabilizing the economy Perhaps the greatest difference in emphasis is the greater reliance placed by the present Administration on a flexible monetary policy, both for restraining an inflation as well as for promoting economic recovery. In the January, 1953 *Report* the following statement was made. "The first objective of credit policy is to assist production, stabilization is an associated purpose . The prime role of credit in production is, of course, not to create demand, but to implement it; its job is to facilitate rather than generate It is possible for credit policy to achieve a victory over inflation by withholding funds from those desiring to buy, although such action may thwart the first objective of the credit mechanism At the other extreme, in a deflationary situation, it is hardly to be expected that credit policy alone can arouse business from recession by making funds available to finance demands which do not exist" (1953, pp. 118, 119)

In contrast, the current *Report* states that the special character of the recent actions of the Federal Government to stimulate the economy "was their promptness and the heavy reliance on monetary policies and tax reductions. . Had it not been for the increased availability of credit and the easing of terms, the fast pace of residential, commercial and state and local construction, which did so much to stabilize the economy during the past year, would not have been attained" (1955, pp 20, 22). The present Administration apparently believes that a flexible monetary policy can generate as well as facilitate demand. Experience with the "active easy" monetary policy of the Federal Reserve Board during the last year appears to substantiate that point of view.

Both *Reports* emphasized the importance of fiscal policy as a stabilizing device, including both tax and expenditure adjustments. The current emphasis on increased spending on public works is due to a desire to promote economic growth rather than to use these expenditures for stabilization purposes. It is contended in the present *Report*, however, that the first line of defense against recession is a strong growth economy Furthermore, the present *Report* argues that although Government can do much to moderate economic fluctuations, there is no reason to believe that they can be completely eliminated One gets the impression from reading the 1953 *Report* that the previous Administration felt that all downward fluctuations in business activity should and could be eliminated Conversely, the present Administration appears more concerned about preventing inflation than was the former group, and strongly recommends use of both monetary and fiscal policies to accomplish this objective

The 1955 *Report* continues the useful earlier practice of using well designed charts for illustrating the written material It is unfortunate that tabular data on the interesting chart on output per man-hour in major industries is omitted (1955, p 5). One extremely useful innovation in the current *Report* is a tabular reconciliation of Federal Government receipts and expenditures from the national income accounts, the consolidated cash budget and the conventional budget for fiscal years 1952–54. As in earlier *Reports*, the current edition presents estimates of the cash and conventional budget for the current and next fiscal years Since most business forecasts are not made on a Federal fiscal year basis, it would be desirable to have these estimates available on a quarterly or calendar basis Also, the inclusion of estimates of future Federal receipts and expenditures on a national income account basis would be useful for forecasting purposes

One of the most significant changes in the use of economic analysis by the two Administrations under consideration is reflected in the changed status of the Council of Economic Advisers. The information provided in the 1954 *Report* on the "Reorganization of the Council" and in the 1955 *Report* on "Activities of the Council" indicates a significant improvement in the effective use of staff advice on economic matters.

In commenting on his tenure as Chairman of the Council of Economic Advisers, Edwin G. Nourse made the following statement: "Causes for the decline of the Council of Economic Advisers are not far to seek. The President [Truman] did not in his initial appointments succeed in finding three economists of the stature needed for the task,

he did not accord them the status in the Executive Office requisite for success; he did not make effective use of them or influence his official family to do so; he did not establish a confidential character for their advisory service or recognize a distinction between economic service and political involvement. . . The general verdict seems to be 'Too much politics and not enough economics So interpreted and so operated, this device of the Employment Act is at best superfluous and at worst mischievous '"[1]

In the 1955 *Report*, light is shed on how the Council is now being used. "In its relation to the President, the Council functions in the economic realm in many respects as the Joint Chiefs of the Staff function in military matters. . . The Council gives its undivided attention to analyzing how the entire economy is faring, to exploring ways and means of adding to its strength, and to advising the President on appropriate economic policies" (1955, p 129).

A further indication of how the Council is being used is found in the following statement: "A representative of the Council, generally the Chairman, reported personally to the President on economic matters once a week, sometimes more often A representative of the Council also appeared regularly at Cabinet meetings to present the Council's thinking about the state of the economy and ways of dealing with the changing economic situation" (1955, p 129) Mr Nourse indicated that "On only half a dozen occasions while I was on the Council were we invited to sit in on a regular or special meeting of Cabinet officers . . . " (p. 384) and then for purposes other than rendering economic advice In addition, at the present time the Chairman of the Council of Economic Advisers is Chairman of an Advisory Board on Economic Growth and Stability made up of representatives of several departments and agencies This group meets weekly and "assures close liaison between the Council and Government agencies that have administrative responsibility for various economic programs. It also provides the Council with timely information and advice on a wide range of current economic issues" (1955, p. 131)

In the absence of more detailed information as to how the Council is now used, it is impossible to compare adequately the functioning of the present Council with the earlier group However, on the basis of the limited information available, it seems extremely likely that the current Council is providing useful economic advice and analysis and that for the first time national policies are being formulated after a careful con-

---

[1] *Economics in the Public Service*, New York Harcourt Brace & Co (1953), p 454

sideration of the economic consequences. If this conclusion is correct, it represents an important step forward in implementing the objectives of the Employment Act of 1946.

The coverage of the statistical tables relating to income, employment, and production is again expanded in the current report. The data are well presented and represent a convenient reference source for researchers in these areas Detailed breakdowns are made when possible and the annual data usually cover a considerable span of years with semiannual, quarterly, or monthly data for more recent years when appropriate.

B *Some Possible Shortcomings*

It is, of course, impossible to analyze adequately in this review each Government policy discussed in the President's *Report* even if the reviewer were competent to do so Therefore, emphasis on possible shortcomings will be placed only on those aspects of the report that appear particularly vulnerable

The 1955 *Report* properly took credit for doing an adequate analysis of the economic situation in early 1954 "The earlier *Report* set forth the conditions of economic progress in our country and in our times By and large, the events of the intervening year have borne out the conclusions of that *Report* concerning the economic state of the Nation and the policies needed to promote sound economic growth" (1955, p 1). The major conclusion on the current outlook of the earlier report was that outlays in most areas would be well maintained in the visible future and the adjustments then in process were not likely to initiate a "cumulative downward movement" of the economy (1954, p. 71)

Although most of the 1954 estimates were of a general rather than specific nature, the explicit estimate of changes in Federal spending, as measured in the gross national product accounts, proved to be extremely poor The 1954 *Report* estimated that by mid-1954 reductions in Federal spending for goods and services might be about $2 billion below the rate at the end of 1953 (January, 1954, p 67) It was stated that "Federal expenditures will continue to be a strong sustaining factor" (1954, p 67). The authors of the 1954 *Report* expected the "small prospective decline in Federal expenditures" to be counteracted in large part by a rise in state and local purchases Actually Federal spending for goods and services in the second quarter, 1954, as measured by the GNP accounts, was down $8 5 billion from the rate in the last quarter of 1953 to a rate of $51.3 billion Thus, the decline was apparently $6.5 billion more in that six-month period than expected The forecast

six-month change was only 23.6% of the actual change or the actual level of Federal spending was 11 3% less than forecasted For the entire year of 1954, Federal spending was $9.8 billion below the fourth quarter 1953 rate, and the rate of State and local spending rose only $1.3 billion.

Private forecasters also generally underestimated the cut in Federal spending in 1954. Yet, in view of the fact the Federal Government has considerable control over this segment of spending, this poor forecast in the 1954 *Report* raises at least two important questions (1) Was this a case of poor liaison among the Treasury, the Council, and the President, or is the gross error in underestimating reductions in Federal spending to be charged to the Treasury? (2) If the Council (or Treasury) had properly estimated the drop in Federal spending, would the Administration have sponsored larger tax cuts in 1954 to assist the transfer of resources from the Federal to the private sector of the economy? Neither of these questions can be answered conclusively on the basis of present information, but they deserve answers by the Administration

On the basis of statements made by Treasury and Council officials during the past two years, it seems probable that the Council was much more concerned about the destabilizing effect of reducing Federal spending relative to taxes during a period of business decline than was the Treasury Treasury officials appeared to be interested primarily in balancing the budget and extending the average maturity of the Federal debt Although these are laudable longer run objectives, they were largely inconsistent with the overall objective of stabilizing the economy in 1954 Fortunately, from a stabilization viewpoint, the Treasury restricted its debt lengthening activities to 6- to 9-year maturities which were bought primarily by commercial banks

In listing six "lessons from experience and guides to the future," the 1955 *Report* states "that contraction may be stopped in its tracks even when Governmental expenditures and budget deficits are declining, provided effective means are taken for building confidence" (1955, p. 22) It cannot be denied that this was accomplished in 1954, but at the expense of reduced economic growth and production, and substantially increased unemployment. It is suggested that a seventh lesson learned (or relearned) from the 1954 experience, is that Federal spending should not be reduced more than taxes during a period of business decline It would seem that the failure to observe this principle in formulating tax policy in 1954 was largely responsible for the decline in 1954, since even the inventory decline was probably substantially due to the

sharp decrease in Federal spending without a compensating rise in the private sector. Although the Administration may be blameless for underestimating the extent of the cut in Federal spending in 1954 due to possible forecasting difficulties, it would seem that the 1955 *Report* is something less than completely frank in failing to emphasize the destabilizing effects of the Federal budget during the past year.

The Administration's case for a higher minimum wage of $0 90 is not convincing and seems somewhat inconsistent with the emphasis on free markets The 1955 *Report* argues that since the present minimum of $0.75 was established "the cost of living and average hourly earnings have risen, providing reason for an increase in the minimum wage when, as at present, the economic outlook is favorable" (1955, p 58). The *Report* points out that minimum wage laws do not get at the fundamental causes of low incomes or poverty and that a higher minimum wage would add appreciably to the costs of certain industries, notably in the South Furthermore, it states: "Nevertheless, $0 90 an hour is the highest minimum wage that can be economically justified in present circumstances A higher minimum might well cause lower production and substantial unemployment in several industries, and—whether directly or indirectly—it would probably bring generally higher prices in its wake" (1955, p. 58)

Chart B-3 on page 90 of the 1955 *Report* shows that unemployment rates tended to be high in the South as of June 30, 1954. Unless very low wage rates in the South are normally associated with monopsonistic conditions, it would seem reasonable to expect that if a $0 90 minimum wage raises wage rates in the South, unemployment rates in this region will be raised To the extent that improving business conditions offset this effect, wages would probably rise in any event. Except in a few small isolated communities, there is little reason to believe that the Southern employer is immune to competitive forces. The effect of a minimum wage in excess of $0 90 would differ from the effects of a minimum wage of $0.90 only in degree, rather than kind. In summary, the Administration's defense of this proposal makes it appear to regard the $0 90 minimum wage recommendation as yielding the maximum political benefit for the moderate amount of economic dislocation allowable

The *Report's* recommendation that the lending authority of the Federal Government for granting loans to small business "should be enlarged so that loans may continue to be made to small concerns that cannot obtain adequate financing on reasonable terms" (1955, p 50) also appears to be inconsistent with the overall emphasis on free mar-

kets. There is, of course, always a shortage of "adequate financing on reasonable terms" when the rate considered to be "reasonable" is substantially below the rate justified by the risks involved. The very sharp increase in the number of new business firms in the postwar period does not indicate substantially increased financial barriers to entry into business. There is little reason to believe that Government officials are better qualified for determining credit worthiness than the local banker or other sources of finance which have detailed information relating to the nature of the prospective borrower. The granting of funds by Government at low rates to finance businesses that are otherwise denied such funds has the effect of giving control over resources to relatively inefficient producers In addition, this program may have a destabilizing effect on business activity as the demand for loans at relatively low rates of interest tends to rise during periods of high business activity and fall during periods of declining business It is difficult to see how this recommendation could have been based primarily on economic analysis

It has been suggested that the present Administration appears to be much more concerned about the evil effects of inflation than about the undesirable effects of reduced business activity and employment Certainly it is fair to say that the present Administration is more concerned about avoiding inflation than were its predecessors Yet, it appears to the reviewer that the present *Report* is correct in emphasizing the desirability of avoiding either extreme and that the policy suggestions have been largely consistent with this goal

On the whole, the current *Report* is well done and is a credit to the Council of Economic Advisers, which assisted in its preparation It should prove of interest and use to all who are interested in the application of economic analysis to the formulation of public policy

STATISTICAL ABSTRACTS

All communications concerning this section should be addressed to the Abstracts Editor, Professor George E. Nicholson, Jr , Chairman of the Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Armitage, P., "A note on the time-homogeneous birth process," *Journal of the Royal Statistical Society*, 15 (1953), 90–91.

Some limiting properties of the time-homogeneous birth process, valid for small exposure times, are obtained Included are an expression for $p_x(t)$, the probability that exactly x events take place in the time interval $(0, T)$, and a demonstration that the mean exposure time remaining in the time interval $(0, T)$ after the xth event has occurred is asymptotically $T/(x+1)$ RICHARD G CORNELL, *Virginia Polytechnic Institute.*

Bailey, N. T. J., "On queueing processes with bulk service," *Journal of the Royal Statistical Society*, Series (B), 16 (1954), 80–87.

This paper investigates mathematically the queueing problem in which a single queue forms from elements arriving at random, order of arrival being preserved, and the queue being then served in groups. The mean and variance for the queue length, and the average waiting time is derived A crude but useful inequality for the average waiting time is established with a suggested application to hospital out-patient clinics. R I TAYLOR, *Virginia Polytechnic Institute.*

Basu, D., "On the optimum character of some estimators used in multistage sampling problems," *Sankhya*, 13 (1954), 363–68.

Three different methods of selecting primary units from strata are discussed primary units chosen with replacement and with different probabilities, primary units chosen without replacement and with equal probabilities, primary units chosen with replacement, and using the first estimate for a primary unit r times if that unit occurs r times. For the three methods, unbiased estimators are simply deduced under general conditions, and these estimates are shown to be the "best" estimates within a class of estimators. For the first and last methods, the problem of selecting the probabilities of selection of given units is studied and optimum theoretical solutions obtained PAUL N. SOMERVILLE, *Virginia Polytechnic Institute.*

Basu, D., and Laha, R. G., "On some characterization of the normal distribution," *Sankhya*, 13 (1954), 359–62.

Geary's theorem is extended by proving that if the sample mean is distributed independently of any $K_r (r \geqq 2)$ (where K_r is the unbiased estimator as given by Fisher of the rth population cumulant K_r) then the parent population is normal T S. RUSSELL, *Virginia Polytechnic Institute.*

Box, G. E. P., "The exploration and exploitation of response surfaces Some general considerations and examples," *Biometrics*, 10 (1954), 16–59.

When the experimenter has several quantitative variables the levels of which he may control in an effort to maximize yield, or minimize cost, classical methods are not particularly apropos The basic idea is that response Y may be representable as a second degree function of the quantitative factors $x_1 x_2 \cdots x_k$ The coefficients can be estimated only if the experimental combinations are suitably chosen The interpretation of the fitted response function is greatly facilitated by reduction of the general quadratic obtained to canonical form in variables $X_1 X_2 \cdots X_k$ where $X_1 \cdots X_k$ are (after choice of origin) obtained by orthogonal transformation of $x_1 x_2 \cdots x_k$ Interpretation of the response function in terms of $X_1 \cdots X_k$ may lead to recognizing factor combinations which provide higher yield, or to identifying factor combinations of equivalent yield but varying cost or convenience.

The principles are illustrated with worked examples including good figures and tables L E. MOSES, *Stanford University*

Box, G. E. P., and Hunter, J. S., "A confidence region for the solution of a set of simultaneous equations with an application to experimental design," *Biometrika*, 41 (1954), 190–99

The authors extend Fieller's theorem to determine an exact confidence region for the solution to a set of k linear equations The confidence region depends on (1) the magnitude of the errors in estimating the coefficients in the equations and (2) the state of the conditioning of the equations A poorly conditioned set of equations is one in which one or more of the equations are almost linearly dependent on other equations An example is presented of a well and a poorly conditioned pair of equations An equation for the boundary of the confidence region is given and illustrated for an empirically determined stationary point on a fitted quadric surface with two independent variables This procedure is generalized to confidence limits for a stationary point on a surface represented by *any* equation linear in the coefficients G I PAUL, *North Carolina State College*

Broadbent, S. R, and **Kendall, David G.**, "The random walk of Trichostrongylus retortaeformis," *Biometrics*, 9 (1953), 460–66

The life cycle of an intestinal parasite of sheep or rabbits, Trichostrongylus retortaeformis, involves a phase in which the larva wanders apparently at random until he climbs a blade of grass, where he remains until eaten The statistical problem treated is to find the distribution of larvae thus trapped on blades of grass It is assumed that the larva's motion until trapping is a random walk of the Brownian movement type, two hypotheses defining the probability of being trapped are considered Both models are analyzed Some empirical data suggest that the Brownian motion assumption may not be justified L E MOSES, *Stanford University*

Bryan, W Ray, "The relative precision of dose response data in the virus and tumor fields as compared with that in certain other fields of biology," *Journal of the National Cancer Institute*, 15 (1954), 305 ff

A commonly used "index of precision" for a biological assay technique is defined as $\lambda=\sigma/\beta$ where σ is the standard deviation of the response metameter around the (straight line) regression of that metameter on log dose, and β is the slope of the regression line The usefulness of the index is commented on A large number of bioassays of many different kinds of biologically active agents were reviewed, from these studies characteristic ranges of values for λ are presented in summary form, classified by type of agent L E MOSES, *Stanford University*

Chanda, K. C, "A note on the consistency and maxima of the roots of likelihood equations," *Biometrika*, 41 (1954), 56–61

The consistency of maximum likelihood estimates for a k-parameter system is proven by use of assumptions which are different from, and claimed to be possibly stronger than, those of Wald R L ANDERSON, *North Carolina State College*

Claringbold, P J, **Biggers, J. D.**, and **Emmens, C. W.**, "The angular transformation in quantal analysis" *Biometrics*, 9 (1953), 467–84.

One of several transformations alternative to the probit transformation is given by $\phi(p)=\sin^{-1}\sqrt{p}$ If for p the observed fraction is used, a quick non-iterative solution is obtained If for p the expected fraction of response is used, an iterative process gives maximum likelihood solution The two methods were both used on several experiments and differences in results were small Little or no information inheres in zero or hundred per cent responses Choice of suitable experimental design will reduce inefficiency from this source A parallelogram design for factorial experiments with quantal response is introduced and illustrated L E MOSES, *Stanford University*

Cochran, William, and **Carroll, Sarah Porter**, "A sampling investigation of the efficiency of weighting inversely as the estimated variance," *Biometrics*, 9 (1953), 447–59.

Let $x_i (i=1, \ldots, k)$ be normally and independently distributed around a common μ, but with different variances σ_i^2 If the σ_i^2 are known then the best estimate of μ is

$$\bar{x}_w = \sum_{i=1}^{k} \frac{w_i x_i}{w}$$

where

$$w_i = \frac{1}{\sigma_i^2}; w = \sum w_i$$

If the σ_i^2 are unknown, but estimates s_i^2 with n_i degrees of freedom are in hand, one might estimate μ by

$$\bar{x}_{\hat{w}} = \sum \frac{\hat{w}_i x_i}{\hat{w}}$$

where

$$\hat{w}_i = \frac{1}{s_i^2} \quad \hat{w} = \sum \hat{w}_i$$

This paper investigates the variance of $\bar{x}_{\hat{w}}$

(it is unbiased) for various values of k and n (taking all $n_i = n$). Sampling methods afford most of the results. The ratio $\sigma_{\hat{\bar{x}}_w}^2/\sigma_{\bar{x}_w}^2$ increases with k. Comparison is made with a formula of Meier for estimating $\sigma_{\hat{\bar{x}}_w}^2$ and an empirical adjustment to it is proposed. L. E. MOSES, *Stanford University*

Cohen, A. C., and Woodward, John, "Tables of Pearson-Lee-Fisher functions of singly truncated normal distributions," *Biometrics*, 9 (1953), 489–97

The estimation of μ and σ in the frequency function

$$f(x) = \frac{1}{I_0\sqrt{2\pi}} e^{-(1/2)((x-\mu)/\sigma)^2} \quad x_0 \leqq x \leqq \infty$$

by the method of maximum likelihood leads to the solution of awkward transcendental equations. Tables are given which eliminate nearly all the computational labor beyond calculating Σx and Σx^2, the sufficient statistics. L. E. MOSES, *Stanford University*

Coombs, C. C., "A method for the study of interstimulus similarity," *Psychometrika*, 19 (1954), 183–94

Various methods for constructing unidimensional scales along a psychological dimension are discussed. The methods that Coombs presents in this article are particularly appropriate for the development of metric scales peculiar to individuals and for testing whether different individuals have the same latent structure for a given set of stimuli. Information provided by several of the methods for collecting data to be scaled is evaluated, the weaknesses and strengths of each of the proposed methods are compared. In addition to evaluating standard methods for scaling, Coombs introduces the method of similarities and the method of cartwheels, the utility of variations of these latter methods is also pointed out. B. J. WINER, *Purdue University*

Dixon, W. J., "Power under normality of several nonparametric tests," *Annals of Mathematical Statistics* 25 (1954), 610–16

The author computed the power of four nonparametric tests (rank-sum, maximum deviation, median and total number of runs) for the difference in means of two (small) samples ($N_1 = N_2 \leqq 5$) drawn from a normal population with equal variance against the alternative $|\mu_1 - \mu_2|/\sigma$ with small level of significance. He used the power efficiency function to compare these with the t-test and he found that the four nonparametric tests have high power efficiencies. The power efficiency decreases slightly for more distant alternatives and as the level of significance increases it increases slightly for the rank sum test while that of the median and maximum deviation tests decrease. The author calculated also the power efficiency for the tests randomized to a single level of significance $\alpha = .025$ to make the comparison simpler. From this he found that the rank sum test has greater power than the median and maximum deviation tests. Furthermore, he calculated the limiting power efficiencies for the rank sum test for $N_1 \leqq N_2 \leqq 5$ and he found that the local power efficiencies for this test is greater than $3/\pi$ which is the limiting local power efficiency for large samples. A. E. SARHAN, *University of North Carolina*

Dunnett, C. W., and Sobel, Milton, "A bivariate generalization of Student's t-distribution with tables for certain special cases," *Biometrika*, 41 (1954), 153–69

The authors consider the simultaneous distribution of two variates, $t_1 = z_1/s$ and $t_2 = z_2/s$. The z_i follow a normal bivariate distribution with zero means, the same variance σ^2, and correlation ρ. The variance σ^2 is assumed independently estimated by s^2 with n degrees of freedom. The probability integral is

$$\text{Prob}\ \{t_1 \leqq h,\ t_2 \leqq h\} = P$$

Tables of P and h are presented for $n = 1\ (1)\ 30\ (3)\ 60\ (15)\ 120, 150, 300, 600$ and ∞, and $\rho = 0.5$ and -0.5. P is given to five decimal places for $h = 0\ (.25)\ 2.50$ and 3.00, plus some additional values for larger h when n is small. h is given to three decimal places for $P = .50, .75, .90, .95$ and $.99$. An asymptotic expansion is derived for P and h,

$$P = \sum_{i=0}^{4} A_i/n^i \quad \text{and} \quad h = \sum_{i=0}^{4} B_i/n^i.$$

Values of the A_i and B_i are presented for the same values of h and P mentioned above. This distribution has applications in certain multiple decision problems. R. L. ANDERSON, *North Carolina State College*

Ghurye, S. G., and Robbins, Herbert, "Two-stage procedures for estimating the difference between means," *Biometrika*, 41 (1954), 146–52

Given two populations, P_i, with unknown means, θ_i, and variances, σ_i^2. Samples of n_i are obtained, at a total cost not to exceed A_0, to estimate $\theta = \theta_1 - \theta_2$. Assuming the cost is $a_1n_1 + a_2n_2 + a_3 \leqq A_0$, the minimum variance estimate is obtained when $n_i^0 = (A/a_i)$

$\sqrt{a_1}\Sigma(\sqrt{a_j}\sigma_j)$, where $A=A_1-a_1$. A two-stage sampling plan is considered when the σ_i must be estimated from initial samples of m_i from P_i. These estimates are used to determine the additional number of observations needed to estimate θ as above. For normal P_i, the variance of this estimate is compared with that obtained when the σ_i are known for $2n_1=2n_2=N=30$ and 50; $m_1=m_2=m(m/N=.2, .3, .4)$, and $\sigma_2/\sigma_1=1.00$ (0.25) 3 00 In no case does the necessity of estimating the σ_i increase the variance by 10% and generally much less. It is also shown that the ratio of the two variances is asymptotically unity for P_i with finite σ_i^2, which also meet some other not very restrictive requirements The usual sampling procedure of assuming $\sigma_1=\sigma_2$ is asymptotically inferior to the two-stage procedure when $\sigma_1<\sigma_2$ E E SAID, *North Carolina State College*

Guest, P. G., "Grouping methods in the fitting of polynomials to equally spaced observations," *Biometrika*, 41 (1954), 62–76

Given n equally spaced observations The first part presents methods of fitting by use of N groups ($n=rN$), including estimated standard errors Relative efficiencies are computed, including the case when some observations must be omitted because n is prime; the loss of efficiency in the latter case may be very serious, especially for high-degree coefficients. A second part is devoted to the use of step-functions By use of double-step functions, high efficiency is obtained for second and third degree polynomials The chief weakness of step-function fitting is the difficulty of estimating standard errors and the degree of polynomial to be used; it is particularly useful when previous experience indicates a linear relationship is satisfactory Comparative time studies are included of the various methods of fitting R L ANDERSON, *North Carolina State College*

Jonckheere, A. R., "A distribution-free k-sample test against ordered alternatives," *Biometrika*, 41 (1954), 133–45

By extending the definition of Kendall's S statistic of rank correlation for two rankings to the case of k rankings, the author furnishes a test of the hypothesis that k samples were obtained from a single population against the alternative that they were drawn from a specific ordering of k populations The exact distribution of S is derived for the general case and is expressed in the form of a recursion formula for the special case in which the samples are of equal size. Tables of Prob ($S\geqq S_0$) are given for certain combinations of $3\leqq k\leqq 6$, $2\leqq m\leqq 5$, and $0\leqq S_0\leqq 96$ for tests in which the k samples are all of size m. The author shows the limiting distribution to be normal when at least two of the k samples increase without bound and gives formulas for approximate tests using the normal and t distributions when samples are large. D. A. GARDINER, *North Carolina State College.*

Mandel, L., "Grading with a gauge subject to random output fluctuations," *Journal of the Royal Statistical Society*, Series B, 16 (1954), 118–30

"A study is made of the errors arising in grading a normal population into classes with a gauge subject to output fluctuations. This involves evaluating the bivariate normal integral when the correlation coefficient is close to unity and series for this are developed Curves are derived showing the dependence of the errors on the population variance and on the instrumental noise level. Means of reducing some errors by small adjustments of the gauging limits are also examined The results have a direct bearing on the industrial applications of β-ray thickness gauges" HALE C. SWEENY, *Virginia Polytechnic Institute.*

Mann, H. B., "A theory of estimation for the fundamental random process and the Ornstein Uhlenbeck process," *Sankhya*, 13, Part 4 (1954), 324–50.

In Chapter 1, the author discusses a random process of the form $y_t=x_t+f(t)$, where x_t is a Fundamental random process and $f(t)$ a function satisfying given assumptions. Maximum likelihood estimates are derived for the variance constant of the Fundamental random process and the parameters of $f(t)$ Together with a discussion of the optimum properties of these estimates a test of significance is provided for testing the hypothesis $f(t)=0$ against specific alternatives In Chapter 2 the process x_t is taken to be an Ornstein Uhlenbeck process depending on two parameters β and σ^2 Methods of estimation are developed for β, σ^2, and the parameters in $f(t)$. Variances and covariances are given for the estimates of the parameters of $f(t)$. JOHN E FREUND, *Virginia Polytechnic Institute.*

Mann, Henry B., and Moranda, Paul B., "On the efficiency of the least square estimates of parameters in the Ornstein Uhlenbeck process," *Sankhya*, 13, Part 4 (1954), 351–58.

Considering a random process of the form

$y_t = x_t + f(t)$, where x_t is an Ornstein Uhlenbeck process and the function $f(t)$ is of a given form and depends on a set of parameters k_i, the limiting form of the maximum likelihood equations for the k_i is derived under the assumption that β, one of the parameters of the Ornstein Uhlenbeck process, is known. The maximum likelihood estimates are subsequently compared with the corresponding least square estimates of the k_i obtained without the assumption that β is known. It is shown that these two sets of estimates are asymptotically of equal efficiency JOHN E. FREUND, *Virginia Polytechnic Institute*

Masuyama, Motosaburo, "Analysis of the 1939 model sample survey results from the viewpoint of integral geometry," *Sankhya*, 13, Part 3 (1954), 229–34

When sampling a geographical region, a frequent method of choosing sub-areas to be sampled is to place a square or grid on a map according to certain prescribed procedures which need not concern us here The sub-area chosen is then the area contained in the grid. The cost of running such a survey is largely dependent on the number of plots or fields contained in the grid since this is the number (p) of fields which must be enumerated. The author discusses a formula for estimating p in terms of the size of grid a^2, the total area covered by the survey in question T, the total area of plots ϕ, the total length of their perimeters λ, and the total number of plots in T, say v The formula is $p = \phi/T + (2\lambda/\pi T)a + (V/T)a^2$ Since the first term on the right of this equation is less than one, we may neglect it and use the formula $p = (2\lambda/\pi T)a + (V/T)a^2$ W. A. THOMPSON, JR., *Virginia Polytechnic Institute.*

Mathen, K. K., and Poti, S. J., "An adjustment for the effect of changing birth rates on infant mortality rates," *Sankhya*, 13 (1954), 417–22.

A theoretical basis is established for estimating weights used in measuring the relative mortality rate in a given year of babies born in that year to those born the previous year It is shown that these weights are the same regardless of whether we assume a linear trend in the birth rate over the two year period or we assume a constant, but not necessarily equal, birth rate for each of the two years. The estimation process is illustrated using data obtained in India, the United Kingdom, and the United States RICHARD G. CORNELL, *Virginia Polytechnic Institute.*

Mukherjee, M., "Estimation of national consumption of the United Kingdom from family budget studies," *Sankhya*, 13 (1954), 412–16.

Estimates of the national consumption pattern from the consumption pattern as given by the family budget inquiries are compared with the official figures of the national consumption pattern. Logical explanations are given for the discrepancies between the estimates and the official figures. T. S. RUSSELL, *Virginia Polytechnic Institute.*

Page, E. S., "An improvement to Wald's approximation for some properties of sequential tests," *Journal of the Royal Statistical Society*, Series B, 16 (1954), 136–39.

"A simple modification of the Wald approximation to the operating characteristic and average sample number of a sequential test is given which provides estimates of these values for starting points of the test near the boundaries and an improved approximation for general starting points."

Wald's approximate formula for the operating characteristic of a sequential test is valid for starting points not near either boundary and for mean paths inclined at not more than a small angle to the boundaries The modification in this paper provides a better approximation R. A. BRADLEY, *Virginia Polytechnic Institute.*

Page, E. S., "Control charts for the mean of a normal population," *Journal of the Royal Statistical Society*, Series (B), 16 (1954), 131–35.

Discussed in this paper is a manner of choosing the sample size and control limit for control charts of the mean of some dimension of a manufactured article. Tables for finding these values for most economical operation are included for the "one-sided" chart (a chart with a single control limit) Tables can also be used for the two-sided case. One set of tables minimizes the average run length under undesirable conditions (more than a given deviation from the mean) for a given average run length under ideal conditions The other set of tables maximizes the average run length under ideal conditions for a given average run length under undesirable conditions R. I. TAYLOR, *Virginia Polytechnic Institute.*

Patterson, H. D., "The errors of lattice sampling," *Journal of the Royal Statistical Society*, Series (B), 16, Part 1 (1954), 140–49.

This paper presents a scheme for using lattices as a pattern for sampling certain types of populations The samples included equal number of units from each main class or from combinations of two, three, or more main classes in different classifications Methods are given for estimation of the sample variances from sample data In general, estimates of errors can be obtained from samples consisting of special patterns of units or from random samples of several mutually exclusive sets each of which itself constitutes a lattice sample The author uses the following lattices in his sampling scheme lattices with "n" groups, square lattices, cubic lattices, and rectangular lattices The case of the square lattice is given in detail and rigorous proofs of the various formulas are given For the other type lattices, in general, simple expressions for the sampling errors are shown BOYD HARSHBARGER, *Virginia Polytechnic Institute.*

Pillai, K C S, and Ramachandran, K V, "On the distribution of the ratio of the ith observation in an ordered sample from a normal population to an independent estimate of the standard deviation," *Annals of Mathematical Statistics*, 25 (1954), 565–72

In problems dealing with the distribution of any order statistic from $N(0, 1)$ population, we run into the powers of the incomplete normal integral in the interval $-\infty$ to x and 0 to x The authors have given an expansion of the normal probability integral and its powers in the interval $-\infty$ to x and 0 to x and have used the results to obtain the distributions of the studentized maximum modulus u_n and the studentized extreme deviate q_n, from the population mean Upper 5 per cent points of the studentized extreme deviate from the population mean and upper and lower 5 per cent points of the u_n are given for small sample sizes and for different d f of the sample standard deviation Upper points of q_n and u_n are useful for simultaneous confidence interval estimation Lower points of u_n are useful for tests of normality against rectangular alternatives A E SARHAN, *University of North Carolina*

Pimentel-Gomes, F., "The use of Mitscherlich's regression law in the analysis of experiments with fertilizers," *Biometrics* 9 (1953), 498–516

Let Y be the yield at a level of fertilizer application X Mitscherlich's Law postulates that $E(Y)=\alpha+\beta\rho^x$ were α, β, and ρ are unknown parameters, and have operational meaning in the context If 4 or 5 equally spaced levels of X are used and a least squares fit is attempted, the resulting equations, though formidable, can be solved non-iteratively by the use of some tables given in the paper Applications of the model are discussed An example is worked L E MOSES, *Stanford University*

Plackett, R. L., "The truncated Poisson distribution," *Biometrics*, 9 (1953), 485–88

The Poisson distribution is defined by $Pr=(\lambda^r e^{-\lambda})/r!$ Cases arise where no observations are available for $r=0$ A method of unbiased estimation of $\theta(\lambda)$ (some function of the parameter) is proposed Where $\theta(\lambda)=\lambda$ the method has high efficiency and advantages in computational ease L E MOSES, *Stanford University*

Read, D. R, "The design of chemical experiments," *Biometrics*, 10 (1954), 1–15

Many experimental programs in chemistry have as their purpose to determine that combination of levels of controllable conditions which will conduce to maximize yield Box and Wilson have given methods for designing and analysing such experiments These methods are discussed in a clear and elementary way and an illustrative example is given L E MOSES, *Stanford University*

Ruben, H, "On the moments of order statistics in samples from normal populations," *Biometrika*, 41 (1954), 200–27

The moments of order statistics derived from normal populations, as well as the moment-generating function of the square of any order statistic, are shown to be closely related to the contents of the members of a class of hyperspherical simplices Formulas for computing the contents of regular hyperspherical simplices and the corresponding moments of the extreme order statistics are derived For samples of up to fifty items, selected surface contents are tabulated plus the first ten moments about the origin and the first four moments about the mean of extreme members to more decimal places than in previous publications The author hopes to compute by electronic means the relative contents of skew hyperspherical simplices and to apply these values for the derivation of at least the first four moments of order statistics ~~which are~~ not extreme T. M KELLEHER, *North Carolina State College*

Rushton, S., "On the confluent hypergeometric function $M(\alpha, y, x)$," *Sankyha*, 13 (1954), 369–76

Rushton, S, and **Lang, E. D**, "Tables of the confluent hypergeometric function," *Sankhya*, 13 (1954), 377–411.

Certain properties of the confluent hypergeometric function

$$M(\alpha, y, x) = 1 + \frac{\alpha}{y}\frac{x}{1!} + \frac{\alpha(\alpha+1)}{y(y+1)}\frac{x^2}{2!} + \frac{\alpha(\alpha+1)(\alpha+2)}{y(y+1)(y+2)}\frac{x^3}{3!} + $$

are stated and reference is made to the different forms in which this function can be written The important Kummer's relation $M(\alpha, y, x) = e^x M(y-\alpha, y, -x)$, some fundamental recurrence formulas and Whittaker's form of the confluent hypergeometric function, $W_{km}(Z)$, is mentioned The application of $M(\alpha, y, x)$ in sequential tests of composite hypotheses in the following cases is indicated (*a*) the one-sided sequential *t*-test, (*b*) the sequential *F*-test, and (*c*) the two-sided sequential *t*-test as a special case of (*b*)

The construction of tables of the confluent hypergeometric function and methods of extending such tables are considered in the paper The tables published give values of $M(\alpha, y, x)$ for $y = 0\ 5, 1\ 0, 1\ 5, 2\ 0, 2\ 5, 3\ 0, 3\ 5$, and $4\ 5$, $x = 02(\ 02) - 10(\ 10) - 1\ 0(1\ 0) - 10$, 20, 30, 50, 100, 200 and a range of half-integer and/or integer values of α from 0 5 to approximately 50 The values of the function are given to 7 significant figures These tables must be regarded as companion tables to those constructed by Nath (*Sankhya* 11 (1951)) giving values of the function for $y=3$ and $y=4$ D E W SCHUMANN, *Virginia Polytechnic Institute*

Som, Ranjan K, "Seasonality in the incidence of strikes in the Bombay textile industry," *Sankhya*, 13 (1954), 423–28

Monthly data from January 1943 to December 1951 of the number of strikes "started in each month" in the textile industry in Bombay state is analyzed with the conclusion that there is a seasonal strike pattern Methods used are "heuristic in nature" First, paired *t*'s for all combinations of months are computed and χ^2 used to test the agreement between the theoretical and observed *t*-distributions Second, using a method attributed to A Wald effects due to trends are eliminated and an analysis of variance of seasonal variations performed PAUL N SOMERVILLE, *Virginia Polytechnic Institute*

Steinhaus, H, "Quality control by sampling (a plea for Bayes' Rule)," *Colloquium Mathematicum*, 2 (1950), 98–108.

The author defends the "outmoded" thesis that in problems of sampling inspection, and the like, it is not unwise to adopt Bayes' rule, more precisely, to behave as though it were known a priori that lot qualities are uniformly distributed from zero to one Three sorts of argument are adduced for this highly controversial thesis According to a line of thought due to J Oderfeld, "modern" procedures involve unrealistic hypothesis for their formal justification no less than does Bayes' rule In certain main problems, behavior based on Bayes' rule differs little, either in its form or consequences, from behavior based on modern procedures Several illustrations suggest that Bayes' rule is more stimulating to the practical resolution of new problems than are "modern" procedures L J SAVAGE, *University of Chicago*

Thurstone, L L., "An analytical method for simple structure," *Psychometrika*, 19 (1954), 173–82

One of the most controversial phases of applied factor analysis concerns the rotation of the reference axes in the configuration of the variables The concept of simple structure was proposed by Thurstone as a guiding principle, but this concept lacked a precise mathematical formulation Attempts have been made to give this concept such formulation, heretofore none of these attempts has led to computationally feasible techniques

In this paper Thurstone reformulates the concept of simple structure in mathematical terms and describes a practically feasible method for achieving simple structure analytically In the past such solutions have been achieved largely by graphical methods The essential feature of Thurstone's analytical solution depends upon a criterion function

$$\phi_p = \sum_j w_{jp} v_{jp}^2$$

where v_{jp} is the scalar product of a vector j with a reference vector p, and w_{jp} are weights assigned to v_{jp} so as to give the criterion function certain optimum properties with respect to a reference vector p The initial location of the reference vector

p is made by inspection; weights are assigned in accordance with a predetermined set of values, such that vectors nearly orthogonal to p are given maximum weights and those vectors having large projections on p are given minimum weights.

Successive adjustments are made in the initial location of the reference vector p until the criterion function ϕ assumes a minimum value Detailed computational procedures for making initial and iterative adjustments are given; a numerical example is presented The adaptation of this rotation method to IBM equipment is said to be in progress. B J WINER, *Purdue University*

Tsao, Chia Kuei, "An extension of Massey's distribution of the maximum deviation between two-sample cumulative step functions," *Annals of Mathematical Statistics*, 25 (1954), 587–592.

If we have two random samples (each has elements arranged in ascending order) of sizes n and m drawn from continuous distributions with cumulative functions $F(x)$ and $G(x)$ respectively, then we can use two statistics d_r and d_r' (where r represents the rth observation) devised by the author to test the hypothesis $F(x)=G(x)$. He derived the distribution of these two statistics under the hypothesis $F(x)=G(x)$ and tabulated their probabilities for $m=2$. He also showed that if $r=m=n$, the distributions of both d_r and d_r' reduce to Massey's distribution and if $r=1$, d_r reduces to a special case of the exceedance problem of Gumbel and von Schelling He illustrated the use of these statistics with an example of censored samples (in life testing). He also showed for situations where the observations below certain ordered observations are missing, or if the observations are available in descending order, that the two statistics D_r and D_r' are to be used and their distributions are identical with those of d_r and d_r'. A. E. SARHAN, *University of North Carolina*.

BOOK REVIEWS

The Economic Report of the President, U. S. Government Printing Office, January 1955. Pp. x, 203. $0.75.

See the article by Beryl Wayne Sprinkel, pp. 240–248 in this issue.

Standard of Living in India and Pakistan, 1931–32 to 1940–41. *R. C Desai.* Bombay: Popular Book Depot. 1953. Pp xvii, 286 Rs. 20.

MORRIS DAVID MORRIS, *University of Washington*

DURING India's colonial period more than a dozen attempts were made to compute the country's national income, the most notable being Rao's estimate for 1931–32. Since Independence the official National Income Committee has worked out estimates for the years 1948–49 through 1950–51. Now R. C. Desai has made another attempt, this one to determine the level of consumer expenditures for the ten year period 1931–32 through 1940–41. His study is not strictly comparable with either of the others Being a study of consumer expenditures, it is merely one step towards the determination of national income. At the same time the area covered is greater, including the whole of what is now India and Pakistan.

Although India is statistically the best served of all underdeveloped areas, the scholar who attempts to use the vast masses of data finds that they crumble in his hands One of the most striking features of the volume is that in his effort to construct his consumer expenditures estimates Desai has explored virtually all the available statistical material. I can recall no other single volume where the limitations and qualifications of the Indian data are more carefully described. While it is desirable to build up national income estimates independently from income, output, and expenditures data, Desai is forced to construct his tables largely on the basis of product estimates. In fact, what he calls consumer expenditures are mainly estimates of current product available for consumer purchase

The book is divided into three parts. The first part develops estimates of production for crops and crop products, livestock products and fish, textiles, and miscellaneous items taken by consumers The second section works out estimates of quantities available for consumption His estimates of quantities available for consumption distinguish between the portions sold on the market and those distributed outside of the market mechanism. His analysis of the subsistence sector is intelligent and more subtle than one customarily finds. Nevertheless, his estimates of "gross village retention" of food cereals are not very different from the traditional assumption. The traditional estimates of retained food crops ran to about 60 per cent. Desai suggests that 59 per cent of total rice production and 49 per cent of total wheat production did not enter the money sector. In this same section he estimates interregional and international movements of products. The third section brings him to grips with the problem of translating physical quantities into money value terms, including the problem of pricing the quantities that do not

reach the market. Here he also estimates the money values of services.

In his last two chapters he sums up the results of his laborious computations, giving us the value of consumer expenditures for his ten year period in current prices and at constant (1938–39) prices Desai concludes that consumer expenditures at current value work out to Rs 82 5 per capita for 1931–32. Rao's national income estimate for the same year is Rs 67 5 per capita which if adjusted to cover the same area as Desai's would probably be about Rs 60 per capita Unless we assume that both private business and government expenditures were negative, which seems highly unlikely, Desai's estimate is much higher than Rao's My own guess is that Desai's estimate is closer to correct.

Desai's figures show that while consumer expenditures rose during the decade, the rise was not sufficient to offset the increase in population Per capita consumer expenditure declined by nearly 5 per cent during the period. He has no data to show whether private capital formation and government expenditures were growing rapidly enough to offset the fall in consumer expenditures, but his guess is that national income per capita was at least not rising

The least satisfying of Dr Desai's judgments stems from his attempt to calculate the reliability of his estimate Using the same technique as Geary used for his work on the national income of Eire, he concludes that his estimate of consumer expenditures might err to the extent of ±7 2 per cent, which seems overly optimistic In fact using his technique it is impossible to tell what the degree of error is Desai is modest enough, however, not to push this judgment

There are many weaknesses in the individual estimates with which readers will quarrel Many of the estimates are purely conventional Nevertheless, his work is one of great merit, an impressive piece of cautious computation in a field that is exasperatingly intractable

---

Workbook in Business Statistics, Third Edition. *Louis F Hampel* Homewood, Illinois Richard D Irwin Inc , 1953 (Pages are unnumbered) $3 50 Paper

GEORGE HORWICH, *Indiana University*

HERE is a collection of 177 exercises for the beginning student in statistics—or rather, descriptive statistics, since very little of statistical inference, in number or quality, is offered The problems are aimed at business and economics students, all of the examples being drawn from that area This is doubtless a limitation in a good course in statistical methodology, but partially compensating is the fact that the data contained in these exercises are often interesting on their own account

The workbook covers the traditional topics: numbers and ratios, charts and tables, investigations (statistical), frequency distributions and measures thereof, index numbers, time series, correlation, the normal curve, reliability and significance Hampel is particularly strong in the first three topics, and altogether adequate in the rest, given the above qualifications and several

more to follow. The problems can be handled rather well without the use of desk calculators, and this feature will appeal to instructors with many students and few machines Some of the exercises are long, others are suited for class discussion A few of the problems duplicate the essence of others, permitting alternate assignment in successive years. This should contribute to the useful life of the book

The reviewer keenly missed problems built around the following measures an index of inequality in connection with the Lorenz curve, the coefficient of variation in its use of comparing distributions of the same variable, but different means (Hampel applies it only to distributions of different variables), Paasche-type index numbers, index numbers of diverse quantities Perhaps only the last is a serious omission, and this attests to Hampel's extensive coverage of descriptive topics One feels, however, that problems of a more purely geometrical formulation would have been a useful addition to the student's interpretive experience For example, exercises might require the student to analyze frequency polygons (e g , to pair ogives with their corresponding non-cumulative curves) or time series in which more or less of the various components are present

The exercises (168, 169, 170) dealing with the normal curve involve empirical distributions which are not quite normal This is unfortunate, since the main relevance of that curve is in sampling theory where the mathematical ideal prevails Hampel has a lot to say about "representative" samples (35, 36, 38), but there is not much reference to, or distinction from, random samples and related issues The analysis of variance is called for in exercises 176 and 177, but the samples in the latter exercise will not pass a test for homoscedasticity (at the 1% level). Moreover, variance analysis would seem to be a doubtful technique to impose on beginning students

What the workbook does, indeed, require is competent problems conveying the basic theory of testing and estimation The use of normal and t-tables in connection with inferences as to population means is probably the easiest way to accomplish this in a first course An alternative method may be through problems concerning the maximum value of a uniform population In this case, the student must have some command of the basic probability laws. But whatever the technique, this reviewer will recommend the use of Hampel's exercises only if the instructor supplements the workbook with testing and estimation problems of his own or someone else's making.

A solutions manual accompanies the workbook The arithmetic appears to be accurate, with one exception, and that is in problem 177 where the sums of squares are incorrectly computed In addition, sample size, instead of degrees of freedom, is used in that exercise in computing variance This is also done in other problems where the standard error of estimate is involved Hampel seems to be aware of what he is doing, since he labels his solutions, "large sample technique " But this is a vague procedure, and no substitute for using degrees of freedom, which are appropriate for large and small samples alike In the answer to problem 36 the author disregards the variability within strata as a consideration in allocating the total sample In problem

174 he assumes the binomial variance to be stabilized in a problem (not fully defined) of determining sample size. There are other questionable answers and assertions, but these are the major ones.

The workbook is, incidentally, in its third edition. The present edition differs from the previous one (1947) in the addition of 27 problems, the replacement of several others, and a modernization of the already mentioned data.

Pictographs and Graphs. *Rudolf Modley and Dyno Lowenstein* New York: Harper and Brothers, 1952. Pp 186. $4.00

KENNETH W. HAEMER, *American Telephone and Telegraph Company*

THIS is a revision and expansion of Dr Modley's earlier book *How to Use Pictorial Statistics* published in 1937 and now out of print. The present book is an excellent reference and source book for anyone who uses pictographs; it should also be a helpful book for anyone who is in doubt about whether to use them or not.

The authors not only describe how to plan and design pictograph charts and diagrams, but, perhaps even more important, explain why and when this method of presentation is suitable and when not Throughout the book, they have included frequent reminders that pictorial presentation is not a substitute for accurate analysis and sharply focused comparisons. They are alert to the three major mistakes in using this method—using pictures to disguise weak data, using them ineptly, and using them for the wrong audience—and carefully point out how to avoid them.

The content is divided into twelve chapters: seven about pictograph methods, two about conventional charts, one about sources and uses of statistical data, and two about production and reproduction methods.

The chapters explaining what pictographs are, how to design them, and how to develop pictograph charts, are the best information on these subjects that this reviewer has seen A fourth chapter, illustrating other pictograph uses, should be a valuable aid in using pictographs for schematic and diagrammatic presentation The usage chapters, on who uses pictographs and where, provide a good coverage of this subject, and include many good examples of pictograph presentation put to work

The section on sources and uses of statistical data is rather brief, and apparently intended for students and other beginners So are the two chapters on "conventional" charts. The first of these, a collection of familiar curve, surface, and bar charts, includes only about half of the standard chart types, but adequately serves the authors' purpose: to demonstrate that many data are not appropriate for pictograph presentation The chapter on "Cheating with Charts"—contributed by a guest author, Frederick Jahnel—is brief but pithy: It points out several basic misuses of graphic presentation—both pictorial and conventional—and shows the serious effect these may have on the reader's interpretation of the data.

The authors may have done the reader one possible disservice in making pictograph presentation seem almost easy. The fact is that this type of chart is harder to do well than a conventional chart If the advice and ideas in this book are followed, it will be of great help in producing successful pictograph charts and diagrams; but no one is going to turn out pictorial graphics as expert as these until he has gained some of the authors' experience and know how.

Statistical Presentation. *John H. Myers* Paterson, N. J.: Littlefield, Adams & Co., 1950. Pp. 68 $.75 Paper

KENNETH W. HAEMER, *American Telephone and Telegraph Company*

THIS title in the Littlefield college outline series is a good synopsis of how to present data in tabular and graphic form.

The author's statement that "This book provides, in simple language, a basic guide to the various devices that have been found helpful in practice" is a good description of its content and style Because it is brief, the book omits many details, and specialists in this field may feel that there are a great many more things that could be said. There are, but the author has managed to include most of the key ideas needed for a general introduction to the subject

Professor Myers makes several valuable points in this small book: in addition to illustrating and discussing the main forms of presentation, he emphasizes the importance of choosing the form of presentation best suited to the purpose This underlines the need for clearly defining what the exact purpose of the presentation is. Failure to do this produces as many poor charts and tables as an insufficient knowledge of presentation methods.

The author points out the difference between charts for presentation and charts for analysis or computation, emphasizes the importance of presenting an accurate picture of the data, and explains the more common types of presentation errors The use of tabular methods for analysis is not mentioned, but the two basic kinds of tabular presentation are summarized and the major details of tabular design reviewed.

The author says nothing about the presentation of statistical information in words. Yet, there are some important things that could be said on this subject, even in a brief treatise such as this Perhaps in the next revised edition of this book, Professor Myers will include some discussion of how to present statistical information in text form.

The publishers describe this book as "a two-color outline " The second color, green, is used somewhat more for decoration than utility, which seems to be a waste In addition to using the color for major topical headings, they might well have used it in the charts and tables for emphasis and clarity.

In general, this book provides a sound introduction to statistical presentation It should be a good supplementary text for students, and a reliable reference for business and professional workers in statistics who are inex-

perienced in presentation or whose experience has been limited to a narrow field.

Biometrika Tables for Statisticians, Volume I *Edited by E S Pearson and H. O Hartley* Cambridge and New York: Cambridge University Press, 1954 $4 50

"A COMPLETE recasting of the two volumes of *Tables for Statisticians and Biometricians* (1914, 1931) has been undertaken by Professor E. S Pearson and Dr H O Hartley," says an announcement of these important tables "Volume I of the new series contains 12 of the most commonly used tables from the earlier volumes, 26 tables published subsequently in *Biometrika*, mostly since 1940, and 16 tables freshly compiled or drawn from other sources The combination represents a selection of tables most often needed by statisticians and experimentalists in the analysis of their data In the case of certain of the more fundamental items rather more extensive tabulation than is normally required has been carried out, so as to provide the accuracy needed by those concerned with mathematical developments

"The 54 tables are preceded by a substantial [102 page] Introduction This gives definitions of the functions tabulated, some account of methods of interpolation, where required, and many illustrations of the use of the tables In the last connection, fuller accounts are provided for the more specialized and more novel tables than for the standard ones whose applications are well known "

The following tables, grouped into 6 classes, make up the volume:

I *Tables of the Normal Probability Function* 1. The integral $P(X)$ and ordinate $Z(X)$ in terms of the standardized deviate X 2 Values of $-\log Q(X) = -\log \{1-P(X)\}$ for large values of X (Extension of Table 1) 3 Values of X for extreme values of Q and P (Extension of Table 4) 4 Values of X in terms of Q and P 5 Values of Z in terms of Q and P 6 Table for probit analysis

II *Basic Tables Derived from the Normal Function* 7 Probability integral of the χ^2-distribution and the cumulative sum of the Poisson distribution 8 Percentage points of the χ^2-distribution 9 Probability integral, $P(t|v)$, of the t-distribution 10 Chart for determining the power function of the t-test. 11 Test for comparisons involving two variances which must be separately estimated 12 Percentage points of the t-distribution 13 Percentage points for the distribution of the correlation coefficient, r, when $\rho=0$ 14 The z-transformation of the correlation coefficient, $z=\tanh^{-1}r$ 15 Charts giving confidence limits for the population correlation coefficient, ρ, given the sample coefficient, r Confidence coefficients 0 95 and 0 99 16 Percentage points of the B-distribution 17. Chart for determining the probability level of the incomplete B-function, $I_x(a, b)$ 18 Percentage points of the F-distribution (variance ratio) 19 Percentage points of the largest variance ratio, $s_{\max.}^2/s_0^2$.

III *Further Tables of Probability Integrals, Percentage Points, etc., of Distributions Derived from the Normal Function* 20 Moment constants of the mean deviation and of the range 21 Percentage points of the distribution of the mean deviation 22 Percentage points of the distribution of the range. 23. Probability integral of the range, W, in normal samples of size n 24 Percentage points of the extreme standardized deviate from population mean, $(x_n-\mu)/\sigma$ or $(\mu-x_1)/\sigma$ 25 Percentage points of the extreme standardized deviate from sample mean, $(x_n-\bar{x})/\sigma$ or $(\bar{x}-x_1)/\sigma$ 26. Percentage points of the extreme studentized deviate from sample mean, $(x_n-\bar{x})/s_\nu$ or $(\bar{x}-x_1)s_\nu$. 27. Mean range in normal samples of size n 28 Mean positions of ranked normal deviates (normal order statistics) 29 Percentage points of the studentized range, $q=(x_n-x_1)s_\nu$ 30 Tables for analysis of variance based on range 31 Percentage points of the ratio, $s_{\max}{}^2/s_{\min}{}^2$ 32 Test for heterogeneity of variance percentage points of M 33 Test for heterogeneity of variance table to facilitate interpolation in Table 32 34 Tests for departure from normality A Percentage points of the distribution of $a=$ (mean deviation)/(standard deviation) B Percentage points of the distribution of $\sqrt{b_1}=m_3/m_2^{3/2}$ C Percentage points of the distribution of $b_2=m_4/m_2^2$ 35. Moments of $s/\sigma=\chi/\sqrt{\nu}$ and factors for determining confidence limits for σ

IV *Tables Relating to Certain Discrete Distributions* 36 Test for the significance of the difference between two Poisson variables 37 Individual terms of certain binomial distributions

$$f(i\,|\,n, p) = \binom{n}{i} p^i(1-p)^{n-i}$$

38 Significance tests in a 2×2 contingency table 39 Individual terms, $e^{-m}m^i/i!$ of the Poisson distribution 40 Confidence limits for the expectation of a Poisson variable 41 Charts providing confidence limits for p in binomial sampling, given a sample fraction c/n Confidence coefficients, 0 95 and 0 99

V *Miscellaneous Tables* (*Pearson type curves, rank correlation, orthogonal polynomials*) 42 Percentage points of Pearson curves, for given β_1, β_2, expressed in standardized measure 43 Chart relating the type of Pearson frequency curve to the values of β_1, β_2 44 Distribution of Spearman's rank correlation coefficient, r_s, in random rankings 45 Distribution of Kendall's rank correlation coefficient, t_k, in random rankings 46 Distribution of the concordance coefficient, W, in random rankings 47 Orthogonal polynomials

VI *Auxiliary Tables* 48 Powers of integers 49 Sums of powers of integers. 50 Squares of integers 51 Factorials of integers, their logarithms, square roots, and their reciprocals 52 Miscellaneous functions of p and $q=1-p$ over the unit range 53 Natural logarithms, $\log_e x$ 54 Useful constants

W A W

Tables of 10^x. National Bureau of Standards Applied Mathematics Series 27. Buckram bound. Pp. 543. $3 50 (Order from Government Printing Office, Washington 25, D. C.).

THE Bureau's announcement of this table reads: "Although there are a number of handy tables of logarithms to 10 or more places, these tables necessitate the use of inverse interpolation for finding the antilogarithm. Thus, a table of antilogarithms is needed. The present volume gives antilogarithms to the base 10, or 10^x, in the form of two tables, a readily interpolable table for 10-decimal accuracy and a basic radix table for 15-figure accuracy. When used in conjunction with logarithmic tables in any extensive computations involving logarithms and antilogarithms, the *Tables of 10^x* will save considerably more labor than will logarithmic tables used alone

"The only similar table is J. Dodson's *Antilogarithmic Canon*, published over 200 years ago. Besides being extremely scarce, Dodson's table is hard to read, the tabular entries are arranged inconveniently, and there are a comparatively large number of errors in it. *Tables of 10^x* contains Table I, 10^x for $x=0(.00001)$ 1.00000 to $10D$, and Table II, a radix table of $10^{\pm n}10^{-p}$, $n=1(1)999$, $p=3, 6, 9, 12, 15$ to $15D$. Although it has the same interval and number of places, Table I alone provides a great improvement over Dodson's table in that it has all known errors corrected, its entries read vertically instead of horizontally, and it omits no digits from any entry. The ease of performing linear interpolation by machine eliminates the need here for differences and proportional parts The fine interval of 10^{-5} in the argument permits determination of the full 10-decimal places by linear interpolation alone with a small 10th place correction that can be done mentally."

M.A L.

Selected Papers in Statistics and Probability. *Abraham Wald* Edited for the Institute of Mathematical Statistics by T W Anderson (Chairman), H Cramer, H. A Freeman, J L Hodges, Jr, E L Lehmann, A M Mood, and C M Stein New York: McGraw-Hill Book Company, Inc , 1955 Pp ix, 702 $8 00

THIS useful and important memorial to Abraham Wald opens with a one-page biography and an eighteen-page survey of Wald's work, unsigned but presumably written by the Editors. Wald's complete bibliography is then given, numbering 104 items (two-thirds of which appeared during the last twelve years before his death (at the age of 48) and the following year. There then follow 51 papers covering "most of Wald's research in statistics and probability except for the work included in the books under his authorship," *Sequential Analysis* and *Statistical Decision Functions* Of the papers, one, "Testing the Difference Between the Means of Two Normal Populations with Unknown Standard Deviations," has not been published previously. One of the papers is in German, 35 are reprinted from the *Annals of Mathematical Statistics*, and 20 have co-authors.

The format and printing, which is by a photographic process, are excellent and the price is commendably low for so large a volume.

W.A.W.

PUBLICATIONS RECEIVED

Beaumont, Ross A., and Ball, Richard W., *Introduction to Modern Algebra and Matrix Theory.* New York: Rinehart and Company, Inc., 1954. $6.00.

Bennett, M. K. *The World's Food: A Study of the Interrelation of World Populations, National Diets, and Food Potentials.* New York. Harper and Brothers, 1954. $4.00.

Blank, David M. *The Volume of Residential Construction, 1889–1950.* New York. National Bureau of Economic Research, 1954. $1.50 Paper.

Bogue, Donald J., and Harris, Dorothy L. *Comparative Population and Urban Research via Multiple Regression and Covariance Analysis* Scripps Foundation for Research in Population Problems and Population Research and Training Center of University of Chicago, 1954. 90 cents. Paper.

Borenstein, Israel. *Capital and Output Trends in Mining Industries, 1870–1948.* Occasional Paper 45. New York National Bureau of Economic Research, 1954 $1 00 Paper

British Ministry of Food. *Domestic Food Consumption and Expenditure, 1952.* London· H M. Stationery Office, 1954. 4s Paper.

Brixey, John C., and Andree, Richard V. *Fundamentals of College Mathematics* New York: Henry Holt and Co , 1954 $5 90.

Chamber of Commerce of the United States. *A Look at Modern Health Insurance.* Washington, D. C.. 1954 $1.50 Paper.

Creamer, Daniel, assisted by **Martin Bernstein.** *Capital and Output Trends in Manufacturing Industries, 1880–1948.* Occasional Paper 41. New York National Bureau of Economic Research, 1954 $1.50 Paper.

Davies, Owen L., ed. *Design and Analysis of Industrial Experiments* New York Hafner Publishing Co., 1954. $10 00.

Davisson, Charles N. *The Marketing of Automotive Parts.* Ann Arbor University of Michigan, School of Business Administration, 1954.

Deane, Phyllis. *Colonial Social Accounting.* Cambridge: University Press, 1954. $9 50.

Dearing, Charles L. *Industrial Pensions.* Washington, D. C.: The Brookings Institution, 1954. Paper.

Enrick, N. L. *Quality Control.* New York The Industrial Press, 1954. $4.00.

Enrick, Norbert Lloyd. *Quality Control Through Statistical Methods.* A Modern Textiles Magazine Handbook Number 3. New York· 1954. Paper.

Flores, Ana Maria. *Metodo y Aplicaciones del Muestro Estadistico* Mexico. Departamento de Muestro, July 1954. Paper.

Fruchter, Benjamin. *Introduction to Factor Analysis.* New York: D. Van Nostrand Co , Inc., 1954. $5.00.

Fryer, H. C. *Elements of Statistics.* New York John Wiley and Sons, 1954. $4 75.

Goldsmith, Raymond W. *The Share of Financial Intermediaries in National Wealth and National Assets, 1900–1949.* Occasional Paper 42. New York National Bureau of Economic Research, 1954. $1 50. Paper.

Hammond, R. J. *Food and Agriculture in Britain, 1939–45.* Stanford: Stanford University Press, 1954 $5.00.

Hegeland, Hugo. *The Multiplier Theory.* Lund C W. K Gleerup, 1954.

Hodgman, Donald R *Soviet Industrial Production, 1928–1951.* Cambridge Harvard University Press, 1954. $5 00.

Householder, Alston S. *Principles of Numerical Analysis* New York: McGraw-Hill Book Co., 1953. $6.00.

Hyman, Herbert H., with William J. Cobb, Jacob J. Feldman, Clyde W. Hart. and Charles Herbert Stember. *Interviewing in Social Research* Chicago University of Chicago, 1954 $8.00

Jaffe, A. J., and Carleton, R. O. *Occupational Mobility in the United States, 1930–1960.* New York: Columbia University Press, 1954 $2 75 Paper.

Jones, Lawrence A., and Durand, David. *Mortgage Lending Experience in Agriculture* Princeton Princeton University Press, 1954. $5 00.

Kendall, M. G. *Exercises in Theoretical Statistics· With Answers and Hints on Solutions* New York. Hafner Publishing Co., 1954. $3 75

Kenyon, Karl W., Scheffer, Victor B., and Chapman, Douglas G. *A Population Study of the Alaska Fur-Seal Herd* Special Scientific Report—Wildlife No. 12. June 1954 Paper.

Key, V. O., Jr. *A Primer of Statistics for Political Scientists* New York. Thomas Y. Crowell Co., 1954. $2.50

Larsson, Tage, and Sjogren, Torsten. *A Methodological, Psychiatric and Statistical Study of a Large Swedish Rural Population.* Copenhagen: Ejner Munksgaard, 1954. Paper.

Lazarsfeld, Paul F., ed. *Mathematical Thinking in the Social Sciences* Glencoe, Illinois The Free Press, 1954. $10 00.

Maverick, Lewis A. *Productivity, A Critique of Current Usage.* Published by the author at Southern Illinois University, Carbondale, Illinois. 1955 50 cents Paper

Maxwell, E. A. *An Analytical Calculus, Vol. III* New York Cambridge University Press, 1954 $2 75

McCloskey, Joseph F., and Trefeth, Florence N, eds *Operations Research for Management* Baltimore The Johns Hopkins Press, 1954 $7 50

Miami, University of, Bureau of Business and Economic Research, Florida *75 Housing Areas Selected Data on Housing, Population and Economic Indicators, 1954.* New York. Housing Securities, Inc (250 Park Ave), 1954 Paper

Milbank Memorial Fund *The Interrelations of Demographic, Economic, and Social Problems in Selected Underdeveloped Areas* New York 1954 $1 00 Paper

Morgenstern, Oskar, ed *Economic Activity Analysis* New York John Wiley and Sons, 1954 $6 75 Lithographed

National Bureau of Standards. *Tables of Sine and Cosine Integrals for Arguments* Applied Mathematics Series 32 Washington, D. C U S Government Printing Office, 1954 $2 25

National Industrial Conference Board. *Prices-Costs-Wages* New York 1954 $1 00 Paper

National Manpower Council *A Policy for Skilled Manpower* New York Columbia University Press, 1954 $4 50

Neisser, Hans, and Modigliani, Franco. *National Income and International Trade* Urbana University of Illinois Press, 1953 $7 50

Nelson, Olaf *Stone Age Economist in the Atomic Age* New York Pageant Press, Inc, 1954 $3 00

Palmer, Gladys, assisted by **Carol P Brainerd** *Labor Mobility in Six Cities, A Report on the Survey Patterns and Factors in Labor Mobility, 1940–1950.* New York Social Science Research Council, 1954 $2 25, paper, $2 75, cloth

Paradine, C G., and Rivett, B. H P. *Statistics for Technologists* New York D. Van Nostrand Co, Inc, 1954 $6 75

Polak, Jacques J. *An International Economic System* Chicago University of Chicago Press, 1954 $4 75.

Roose, Kenneth D. *The Economics of Recession and Revival. An Interpretation of the Year 1937–38* New Haven Yale University Press, 1954 $4 00

Sanderson, Fred H *Methods of Crop Forecasting* Cambridge Harvard University Press, 1954 $5 00.

Savage, Leonard J. *The Foundations of Statistics* New York John Wiley and Sons, 1954 $6 00

Smith, A B. *Biomathematics* The Third Edition and a rewritten version of the work by the late W M Feldman New York Hafner Publishing Co, 1954 $13 50

Thomas, Brindley *Migration and Economic Growth* New York Cambridge University Press, 1954 $8.00.

Tuck, R. H *An Essay on the Economic Theory of Rank* Oxford Basil Blackwell, 1954 7s 6d Paper

Ullmer, Melville J *Trends and Cycles in Capital Formation by United States Railroads, 1870–1950* Occasional Paper 43. New York National Bureau of Economic Research, 1954 $1 50 Paper

UNESCO *Basic Facts and Figures* New York Columbia University Press, 1954 $1 00 Paper

U. N Department of Economic Affairs. *Commodity Trade and Economic Development* New York Columbia University Press, 1953 75 cents Paper

U. N. Department of Economic Affairs. *The International Flow of Private Capital, 1946–52* New York Columbia University Press, 1954 40 cents Paper

U S Department of Social Affairs, Population Division. *International Research on Migration* New York Columbia University Press, 1953 30 cents Paper

Vand, Vladimir, and Pepinsky, Ray. *The Statistical Approach to X-Ray Structure Analysis* State College, Pa Pennsylvania State University Department of Physics, 1953 Paper

Wessman, Lucile Tomlinson, ed "*Those Having Torches* " *Economic Essays in Honor of Alzada Comstock* South Hadley, Mass Mount Holyoke College Department of Economics and Sociology, 1954 $3 00

Williams, John Burr *International Trade Under Flexible Exchange Rates* Amsterdam North-Holland Publishing Co, 1954 $5 00

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 270 JUNE 1955 *Volume 50*

A REVIEW OF THE STATISTICAL EVIDENCE ON THE ASSOCIATION BETWEEN SMOKING AND LUNG CANCER

SIDNEY J. CUTLER

National Cancer Institute, United States Public Health Service

Leading investigators generally agree that a significant part of the observed increase in lung cancer mortality represents a real increase in the rate at which lung cancer is developing in the population. This increase cannot reasonably be attributed to genetic change in the human population and therefore must be due to environmental factors. Available evidence linking tobacco smoking to lung cancer is fairly extensive and impressive (1) The increase in lung cancer mortality has been generally parallel to an increase in cigarette consumption, (2) In each of 14 case history studies there was a smaller percentage of non-smokers and a higher percentage of heavy smokers among lung cancer patients than among comparable controls; (3) Preliminary results of two population studies indicate higher mortality from lung cancer among smokers than among non-smokers and a still higher mortality among heavy smokers, and (4) At least one team of investigators has produced skin cancer in animals with condensates of tobacco smoke.

There is disagreement whether the evidence at hand warrants a conclusion that smoking and lung cancer are causally related. The relative importance of smoking, air pollution, and occupational exposure to cancerigenic materials remains to be established

DURING the first half of the 20th century lung cancer emerged from relative obscurity to become an important cause of death in at least a dozen countries [12] The magnitude of the increase has been most unusual for a so-called chronic disease. During the last two decades, alone, the rate of lung cancer mortality in the United States increased by 400 per cent. In 1930 less than 3,000 deaths in the United States were ascribed to lung cancer; preliminary statistics for 1953 attribute about 23,000 deaths to this disease. More men now die from lung cancer than from cancer of any other site.

Part of the reported increase is undoubtedly due to improved diagnostic techniques and to greater alertness on the part of physicians. Some investigators believe that the observed increase in lung cancer mortality is fictitious and due entirely to improved case finding. This position is incompatible with the following facts:

1. The relative increase in males has been much greater than in females. Between 1930 and 1950 the increase in lung cancer mortality in the United States was more than four times as great in males as in females. It seems unreasonable to assume that improved diagnostic techniques have been applied to a much greater extent to men than to women.

2. The relative increase was greater in old than in young people, especially among men For example, the increase in lung cancer mortality was more than five times as large for white men aged 75–84 as for those between 35 and 44 years of age. It doesn't seem reasonable that the quality of diagnosis is materially affected by the age of the patient.

3. Lung cancer mortality is continuing to increase without benefit of any marked improvement in diagnostic techniques in recent years.

The consensus among leading investigators is that a significant part of the observed increase is absolute and represents a real increase in the rate at which lung cancer is developing in the population [3] [1]

The lung cancer death rate in males is now about 4.4 times the rate in females. Prior to age 40, lung cancer is rare in persons of both sexes and occurs with increasing frequency during late adult life and old age. However, the manner in which the lung cancer mortality rate increases with age is markedly different for the two sexes Among males, the mortality curve rises very rapidly between ages 40 and 70 and then declines almost as rapidly In contrast, the curve for females resembles that for all forms of cancer combined, showing a slower but steady rise over the entire life span

Lung cancer mortality rates for the two sexes did not always differ so much in the past. In 1914,[2] the rates for males and females were at

[1] Comprehensive data on the incidence of cancer are scarce and available information covers a relatively short period. One of the most extensive sets of data is provided by the studies of the National Cancer Institute Ten metropolitan areas of the United States were surveyed in 1937–39 and resurveyed in 1947–48. The data pertaining to the incidence of lung cancer are in line with mortality data for the United States. During the period between the two surveys, the incidence of lung cancer increased by 119 per cent among males and by 67 per cent among females (based on age-adjusted rates) During the latter period, lung cancer occurred four and one half times as frequently in males as in females, compared to a male-female ratio of 3.4 during the earlier period.

[2] Deaths due to cancer of the lung were not identified separately in United States mortality statistics until 1914 and were not routinely tabulated until 1930

about the same level and the curve of age-specific rates for each sex was characterized by a slow rise from early adult life until extreme old age. Since then lung cancer mortality rates have increased much more rapidly for males than for females and the shape of the mortality curve for males has changed. The rates for males no longer increase throughout the lifespan, as do the rates for most forms of cancer, they now drop off sharply from about age 70 [11]

The age-specific rates for each year are based on deaths among persons of different ages and hence on deaths among persons born at different times. One may ask, however, what has been the mortality experience of cohorts of people born at about the same time; such as people born around 1890, 1900, 1910, etc. Dorn [11] examined the trend of lung cancer mortality in the United States for cohorts of men and women born from 1850 on. He concluded that the observed trend is consistent with the hypothesis that some time in the past a carcinogenic agent (or agents) capable of producing cancer of the lung had been introduced into the environment and that (a) males were either exposed to it more intensely or for longer periods, or were more susceptible to it, and (b) exposure, at least initially, was greater for young men than for those at more advanced ages This hypothesis serves to explain the observed change in the shape of the curve of age-specific lung cancer mortality rates for males. Clemmesen, Nielsen, and Jensen [4] in examining the trend of lung cancer mortality in Denmark concluded that the observed increase must have been caused by a carcinogenic influence introduced during the early part of the 20th century.[3]

It is interesting to note that, if the evidence pointing to a real increase in lung cancer mortality is accepted, one is led to the hypothesis that it is probably due to environmental factors It would be difficult to attribute an increase of the magnitude observed to genetic change. A number of environmental changes, not mutually exclusive, have been suggested as causal factors in the increase of lung cancer mortality. These are: (1) increased use of cigarettes, (2) increased atmospheric pollution by motor vehicle exhausts, factory wastes, etc., and (3) increased occupational exposure to known cancer producing substances.

No attempt will be made here to evaluate the relative importance of these suspected environmental agents, other than to point out a few observations. The number of persons engaged in occupations involving exposure to known carcinogens is very small. Thus, while additional occupational carcinogens may be identified in the future, only a small part of the observed increase in lung cancer mortality may be at-

[3] The cohort approach was first applied to lung cancer mortality data by Korteweg [16].

tributed to known industrial hazards. As indicated by Heller [14], "the atmospheric pollution theory relies mainly on urban-rural differentials in mortality, plus the demonstration in urban atmosphere of substances which are carcinogenic to animals . . . Inability to classify the population by degree of exposure to atmospheric pollution, and to compute attack rates for groups of people with varying exposure has been a major deterrent in testing the hypothesis of atmospheric pollution as a cause of this disease."[4] Available evidence linking tobacco smoking to lung cancer is fairly extensive and impressive

One reason for suspecting an association between smoking and lung cancer is the observed parallelism between the trends for cigarette consumption and lung cancer mortality Annual per capita consumption of cigarettes in the United States was less than 100 at the beginning of the century, about 600 in 1920, and more than 3,000 in 1950 The parallelism between per capita consumption of cigarettes and the rate of lung cancer mortality has been noted in a number of European countries, as well as in the United States [8]. The existence of a concomitant increase, while necessary, is not sufficient in itself to establish causality. However, the temporal association between cigarette smoking and lung cancer mortality plus the observation by various clinicians that most patients with lung cancer seemed to be heavy smokers led to the hypothesis that smoking is a cause of lung cancer

CASE HISTORY STUDIES

The assertion that there is an association between smoking and lung cancer is based in large part on a series of studies in which the smoking histories of patients with cancer of the lung have been compared with smoking histories of persons free of lung cancer Fourteen studies reported between 1939 and 1954 are described in Table 1 Although a few women were included in several studies, the results pertain essentially to white males The findings are summarized in Table 2. In each study there was a smaller percentage of non-smokers and a higher percentage of heavy smokers among lung cancer patients than among the controls Insofar as is generally known, every study of this type has shown this relationship

Cornfield [6] has shown that data of this type may be used to estimate the relative risk of developing a disease for persons possessing a given characteristic compared with persons not possessing that characteristic. Estimates of the relative risk of developing lung cancer for all smokers

[4] The evidence linking occupational exposure and air pollution to lung cancer has been reviewed by Hueper [15].

and for heavy smokers, as compared to non-smokers, are given in Table 2. All the studies point in the same direction—a greater risk for smokers than for non-smokers, and a still greater risk for heavy smokers However, the estimates of relative risk which these studies yield vary over a very wide range—from 1 2 to 36 4 for all smokers, and from 1 9 to 79 0 for heavy smokers

The variation in results is due in part to differences in defining the various smoking classes For example, Doll and Hill based their classification on the *most recent amount smoked*, whereas Sadowsky, Gilliam, and Cornfield used the patient's *earliest smoking habits* [5] In some cases, cigars and pipe tobacco were converted to an equivalent number of cigarettes, in others the percentage of heavy smokers given in Table 2 pertains to cigarette smokers only. Several of the studies indicate that the association between smoking and lung cancer is different for cigarette, cigar, and pipe smokers Sadowsky, Gilliam, and Cornfield classified their cases as smokers of cigarettes only, cigars only, pipes only, and mixed forms of tobacco The relative risk of developing lung cancer for each of these groups as compared to non-smokers was 4 5, 3 0, 1 2, and 4 5 to 1 respectively Thus, differences in the proportion of cigarette, cigar, and pipe smokers in the various studies may have contributed to the variation in results In addition, part of the variation in results may be attributed to differences in the age distributions of the cases included in the various studies Smoking habits differ among persons in different age classes For example, in Levin's study the percentage of heavy cigarette smokers among the controls decreased from a high of 46 among men 30–39 years of age to a low of 7 among men 70–79 years of age In addition to differences in age between studies, there were differences in the age distributions of the lung cancer and control cases of a particular study Some of the investigators adjusted for age, while others did not

The case history studies summarized above were subject to several possible sources of error—memory error on the part of the respondent, bias on the part of the interviewer or the respondent due to knowledge of the diagnosis, and inadequate sampling techniques

Marketing studies have indicated that, due to memory error, the personal interview and written questionnaire tend to yield inaccurate information concerning an individual's buying habits or consumption of specific products It is, therefore, possible that these clinical studies

---

[5] Doll and Hill used two additional bases for grouping their cases into smoking classes maximum amount of regular smoking, and estimated life-time consumption of tobacco The results for all three methods of classification were very similar

TABLE 1

DESCRIPTION OF 14 RETROSPECTIVE STUDIES OF SMOKING HABITS

Study reported by	Country	Year reported	Lung cancer cases		Controls		Method of collecting data
			No	Description	No.	Description	
Müller [21]	Ger.	1939	86	Hospitalized lung cancer patients in Köln, 1928–39.	86	Sample of men in general population falling within same age group as the patients.	Questionnaire to relatives of cancer patients. Methods used for controls not stated.
Schairer & Schöniger [23]	Ger.	1943	93	Deaths due to lung cancer in Thuringen, 1930–41. Average age was 53 9	270	Sample of men in general population of same area, aged 53–54.	Questionnaire to relatives of cancer cases and to sample of men in general population.
Wassink [25]	Neth.	1948	136	Hospitalized lung cancer patients	100	Sample of men in general population, in same occupational and age classes	Interview Diagnosis known.
Schrek et al. [24]	U S.	1950	82	Lung cancer patients seen in V.A hospital in Illinois, 1942–44	522	Patients with cancer other than of lip, tongue, mouth, larynx, pharynx, esophagus, stomach, and lung admitted during same period.	Interview. Diagnosis known.
Mills & Porter [20]	U S.	1950	444	Deaths due to respiratory cancer in Cincinnati, O., 1940–45 and in Detroit, Mich., 1942–46	430	0 9% sample of white males, 20 years old or over, in Columbus, O in 1947 Adjustment made for difference in age distributions of cases and controls	Mail questionnaire to next of kin of cancer cases. Interviewed controls.
Wynder & Graham [28]	U S	1950	605	Hospitalized lung cancer cases in different parts of the United States	780	Admissions to general medical and surgical service in 3 St Louis hospitals. Adjustment made for difference in age distributions of cases and controls	Interview. Diagnosis known, except for 100 lung cancer cases in a series of 286 admissions with variety of chest ailments.
McConnel et al [19]	Eng	1952	93	Hospitalized lung cancer cases in Liverpool area seen 1946–49.	186	Hospitalized patients without cancer seen in same hospitals 1948–50.	Interview. Diagnosis known.
Doll & Hill [10]	Eng.	1952	1357	Hospitalized cases of lung cancer in various parts of England, 1948–51.	1357	Hospitalized patients without cancer matched for sex, age, hospital and admission date	Interview. Diagnosis known, but in more than 200 cases initial diagnosis of lung cancer was found to be incorrect.

TABLE 1—*Continued*

Study reported by	Country	Year reported	Lung cancer cases		Controls		Method of collecting data
			No	Description	No	Description	
Wynder & Cornfield [27]	U. S.	1953	63	Physicians who died of lung cancer, 1949–52	133	Physicians who died of other forms of cancer, 1950–52	Mail questionnaire to next of kin Diagnosis known.
Sadowsky et al [22]	U S.	1953	477	Hospitalized cases of lung cancer in different parts of the United States, seen 1938–43	615	Patients without cancer seen in same hospitals during same period Adjustment made for difference in age distributions of cases and controls	Interview—smoking data included with histories on a variety of subjects. Diagnosis known.
Koulumies [17]	Fin.	1953	712	Cases examined at a radiotherapy institute during 16-year period, found to have lung cancer	300	Admissions to outpatient department in 1952—men 40 years old or over, no cancer suspected	Interview at time of admission.
Breslow et al [2]	U S	1954	518	Hospitalized cases of lung cancer seen 1949–52 in various parts of California	518	Admissions to same hospitals—matched for sex, age and race; without cancer or chest ailment	Interview Diagnosis known.
Levin [18]	U S	1954	490	Admissions from 1938 on to State cancer hospital in New York, found to have lung cancer	2365	Admissions during same period, found not to have cancer. Adjustment made for difference in age distributions of cases and controls	Interview—smoking data part of regular case history upon admission
Watson & Conte [26]	U S.	1954	265	Admissions 1950–52 to a thoracic clinic in New York City, found to have lung cancer	287	Admissions to same clinic, found not to have lung cancer.	Interview at time of admission.

Note These studies deal almost exclusively with white males

do not provide accurate information on the amount of tobacco smoked. However, to invalidate the results of these studies on the basis of memory error, it would be necessary to assume that memory error was responsible for a systematic overstatement of smoking habits by patients with lung cancer or a systematic understatement by the control groups.

TABLE 2

SUMMARY OF FINDINGS REPORTED IN 14 RETROSPECTIVE STUDIES OF SMOKING HABITS

Study reported by	Per cent non-smokers		Per cent heavy smokers		Relative risk of developing lung cancer. Ratio to non-smokers	
	Lung cancer cases	Controls	Lung cancer cases	Controls	All smokers	Heavy smokers
Muller	3.5	16.3	50.0	10.5	5.4	22.2
Schairer & Schoniger[a]	3.2	15.9	31.2	9.3	5.7	16.7
Wassink[b]	5.0	19.0	55.0	19.0	4.5	11.0
Schrek, et al.[c]	14.6	23.9	18.3	9.2	1.8	3.3
Mills & Porter	7.0	31.0	—	—	6.0	—
Wynder & Graham	1.3	14.6	51.2	19.1	13.0	30.1
McConnel, et al.	5.4	6.5	38.5	23.8	1.2	1.9
Doll & Hill	0.5	4.5	25.0	13.4	9.4	16.8
Wynder & Cornfield	4.1	20.6	67.6	29.3	6.1	11.6
Sadowsky, et al.[d]	3.8	13.2	46.8	30.7	3.8	5.3
Koulumies	0.6	18.0	65.8	25.0	36.4	79.0
Breslow, et al.	3.7	10.8	75.6	44.2	3.2	5.0
Levin[e]	8.0	26.9	54.8	28.8	4.2	6.4
Watson & Conte	1.9	9.7	73.0	57.0	5.5	6.5

Note. Relative risk was computed by means of a technique developed by Cornfield [6, 7].

Heavy smokers are defined here as persons smoking more than one pack of cigarettes per day, or its equivalent. Approximated from variety of smoking classes, with lower limits ranging from 20 to 26 cigarettes per day.

[a] Also compared cases with cancers of the tongue, esophagus, stomach, colon and prostate with the same controls. The stomach cancer cases (128) were found to resemble the controls very closely. The cases with all other forms of cancer combined (98) had fewer non-smokers and more heavy smokers than the controls, but the difference was less striking than for the lung cancer cases.

[b] Numerical values of results approximated from graphic presentation.

[c] Also compared cases with cancers of lip, tongue, mouth, larynx and pharynx, esophagus and stomach to same series of controls. Found no significant differences except for lip and for larynx and pharynx combined.

[d] Also compared cases with cancers of lip, tongue, mouth, pharynx, larynx, esophagus, and skin to same series of controls. Found positive associations between laryngeal cancer and cigarette smoking and between lip cancer and pipe smoking. Found a negative association between smoking and skin cancer.

[e] Also compared cases with cancers of lip, pharynx, esophagus, colon, and rectum to same series of controls. Found no significant differences except for lip.

The studies of Wynder and Graham, Doll and Hill, and Levin indicate that prior knowledge of the diagnosis on the part of the interviewer or the patient did not bias the results In the Wynder and Graham study, 286 men with chest ailments were interviewed prior to diagnosis One hundred were later found to have lung cancer, and 186 were found to have other conditions The smoking habits of these 100 lung cancer cases were very similar to those of the 505 lung cancer cases whose diagnoses were known at time of interview, and the smoking habits of the 186 cases that were found not to have lung cancer strongly resembled those of the 780 non-cancer cases in the control series. In the study of Doll and Hill, and in Levin's, some of the cases believed to have lung cancer at time of interview were later found to have other chest ailments—about 14 per cent and 19 per cent respectively. The smoking histories of these initially incorrectly diagnosed cases were found to be significantly different from those of the correctly diagnosed lung cancer cases and were very similar to those of patients with diseases other than cancer

It has been argued that the use of hospital populations for the investigation of possible association between a disease and a population characteristic may lead to spurious correlation due to uncontrolled factors involved in bringing members of the population to the hospital [1] In this instance, the question is whether an individual's smoking habits may influence the likelihood of his being hospitalized for a particular disease Is a smoker with lung cancer more likely to be hospitalized than a non-smoker with lung cancer, while a smoker with another form of cancer or another disease is less likely to be hospitalized? It seems rather far-fetched to insist that this type of selection operates However, in order to avoid any possible bias which might be involved in the use of hospitalized cases, Wynder and Cornfield based their study on death notices in the Journal of the American Medical Association Questionnaires were sent to the families of physicians who had died of various forms of cancer Physicians were selected because they were believed to represent a population group which is homogeneous economically, with little occupational exposure to respiratory irritants, and with equal access to diagnostic facilities Schairer and Schoniger, and Mills and Porter also obtained their lung cancer cases from death notices. Müller, Schairer and Schöniger, Wassink, and Mills and Porter used samples of the general population as controls These studies indicate that the reported association between smoking and lung cancer is not peculiar to hospital populations

The sampling techniques used in these case history studies were

generally not sophisticated. It may be that the lung cancer cases studied were not representative of all persons with lung cancer and that the controls were not representative of the general population, but these two requirements are not essential. To study the relationship between smoking and lung cancer it is sufficient that the lung cancer cases and controls be drawn from the same population. In two studies (Doll and Hill, and Levin) segments of the lung cancer series and of the controls were definitely drawn from the same population—patients with a presumptive diagnosis of lung cancer at interview. Subdivision of these series into lung cancer cases and non-cancer cases on the basis of final diagnosis yielded fewer non-smokers and more heavy smokers among the patients with lung cancer

Levin [18] has summarized various questions regarding possible sampling error as follows:

"1. How likely is it that the control and lung cancer cases in the studies reported were drawn from two different populations which differed significantly with respect to smoking and differed always so that the control cases came from a lighter-smoking population than the lung cancer cases?

"2 How likely is it that these sampling biases would be more marked for cigarette smokers than for pipe and cigar smokers and more marked for heavy cigarette smokers than for light cigarette smokers?

"3. How likely is it that lung cancer cases who smoke are more apt to go to the hospital or that patients with other diseases who smoke are less apt to do so?

"4 How likely is it that these sampling biases should occur repeatedly in studies made in different countries and in different sections of the same country?"

He concluded that, "None of these contingencies seems very likely, and their occurrence to so marked a degree as to show the difference between smokers and non-smokers elicited by the various studies is even less likely."

Although the association between smoking and lung cancer indicated by the studies discussed above is probably valid for the specific population groups studied—selected segments of white men in the United States and Northern Europe—it may be invalid to generalize to broad population groups, including women, Negroes, Asiatics, etc However, considering that this association has been found in a variety of groups, it would be surprising if it were not found to be widespread Several large scale studies are now under way of the incidence of lung cancer in populations whose smoking characteristics are known.

POPULATION STUDIES

The conclusion from the results of the case history studies, that smokers are more likely to develop lung cancer than non-smokers, is supported by preliminary findings of two large scale population studies—one in the United States and one in England These studies were undertaken to obtain a direct measure of the risk of developing lung cancer for various classes of smokers in specific population groups. Some investigators feel that case history studies fall short as a basis for generalization concerning the relationship between smoking and lung cancer, because it is not known what specific populations generate the lung cancer and control series. In addition, the case history studies reported to date have yielded a wide range of estimates of the relative risk of developing lung cancer for smokers as compared to non-smokers. By determining the incidence of lung cancer in a population whose smoking habits are known, the possible biases resulting from the study of populations of sick people can be avoided Furthermore, by studying large population groups, fairly stable estimates of the lung cancer risk for various classes of smokers can be obtained

Beginning in January 1952, the American Cancer Society obtained smoking histories on approximately 200,000 men, 50 to 70 years of age, in nine different states Volunteer workers each collected information from about 20 men of their acquaintance. Each volunteer reports annually on the survival status of the members of her group. If a man dies, a copy of the death certificate is obtained and in the case of a cancer death, additional information is secured from the appropriate hospital and physician Hammond and Horn recently reported that in 20 months, 4,854 men died, 167 of lung cancer [13] The computed lung cancer death rate was 27 per 100,000 for non-smokers, 113 for men smoking less than a pack of cigarettes a day, and 239 for men smoking one pack or more a day.[6] Thus, on the basis of these preliminary results, men smoking more than a pack of cigarettes a day are about nine times as likely to develop lung cancer as men who do not smoke

In October 1951, Doll and Hill sent questionnaires to all physicians on the Medical Register of the United Kingdom [9] Out of 59,600 doctors, 41,024 replied; 40,564 provided useable information. As doctors die, information on cause of death is obtained from the Registrars-General. The preliminary report is based on 789 deaths in 29 months among male physicians who were 35 years of age or over when they completed the questionnaire; 36 deaths were due to lung cancer The

[6] These are not annual rates.

age-adjusted annual lung cancer death rates per 1,000 population were:

Smoking class	*Rate*
Non-smokers	0 00
1–14 cigarettes per day	0 48
15–24 cigarettes per day	0 67
25 or more cigarettes per day	1 14

As indicated in the case history studies, the risk increases as the amount of smoking increases.

A third large scale population study is being conducted by the National Cancer Institute of the United States Public Health Service in cooperation with the Veterans Administration Early in 1954, questionnaires requesting fairly detailed smoking histories were sent to 291,000 veterans (mainly of World War I) holding U S Government Life Insurance policies Replies have been received from about 220,000 As policies are paid, copies of the death certificates are obtained No results are as yet available

The three studies discussed above might be criticized, because probability sampling was not used It might be contended that it will not be possible to use the findings as a basis for generalization to broad population groups

Whereas probability sampling is desirable, it is frequently impractical Furthermore, as indicated by Cochran, Mosteller, and Tukey [5], there are situations in which sampling may not be "good policy " "The inquirer may not be able to 'afford' the cost in time or money for a probability sample . . The statement 'he didn't use a probability sample' is thus not a criticism which should end further discussion and doom the inquiry to the cellar " Smoking histories can be obtained from the members of a representative sample of the United States population. However, the cost of keeping the members of such a sample under observation for an extended period in order to determine the incidence of lung cancer would be prohibitive One of the virtues of the three population studies now in progress is that the individuals under study can be "followed" at a relatively small cost

No one is likely to insist that no useful knowledge can be obtained in the absence of probability sampling If a consistent relationship between smoking and lung cancer is found in a variety of population groups, the hypothesis that smoking is associated with lung cancer will be greatly strengthened The studies now under way will make it possible to test the validity of this hypothesis for different geographic areas, rural as well as urban populations, and different occupational groups However, since none of these studies include a large number of women, an investigation of the incidence of lung cancer in various classes of smokers in a female population seems to be called for.

DISCUSSION

The demonstration that lung cancer occurs more frequently in smokers and still more frequently in heavy smokers does not and cannot prove that smoking *per se* causes cancer of the lung. In studies of human populations, the individual's past, current, and future smoking practices must be taken as found, they cannot be assigned in advance Whether or not an individual smokes is largely a matter of personal choice. That choice may be bound up with other factors—social, occupational, economic, and biologic—which may be related to predisposition to lung cancer However, if a substantial reduction in smoking were followed, some years later, by a marked decrease in the incidence of lung cancer, this might be considered very strong supporting evidence that smoking "causes" lung cancer.

Some investigators believe that the available evidence is sufficiently strong to warrant the presumption that the relationship between smoking and lung cancer is causal Others want to wait for the final results of the population studies now under way Still others contend that the available evidence is far from conclusive Some are dissatisfied with the nature of the samples in both the case history and population studies in that the members of the various smoking classes are self-selected Attempts are presently being made to investigate the biologic and personality characteristics of heavy smokers, light smokers, and non-smokers. The adherents to the industrial exposure—air pollution hypothesis believe that among the various possible agents, tobacco smoke is only a minor factor in the rapid increase in lung cancer mortality Their position is based in large part on the experimental evidence in animals regarding the carcinogenicity of certain industrial products, and the general lack of success to date in producing cancer in animals with tobacco smoke or its condensates

Since lung cancer occurs among non-smokers, it is evident that if smoking is a cause it is not the only cause. Furthermore, it is difficult to explain observed variation in reported lung cancer mortality in different population groups entirely on the basis of variation in smoking habits For example, the lung cancer mortality rate is about twice as high in England as in the United States, it is roughly twice as high in urban as in rural areas, and it varies over a wide range among urban areas in the United States It seems likely that part of the variation may be due to differences in the quality of diagnosis and in the accuracy of reported causes of death. Some of the variation may be caused by differences in air pollution and industrial exposure How much of the variation may be attributed to smoking in itself or to the co-carcinogenic action of smoking, air pollution, and industrial exposure is not known. Available evidence concerning the association

between smoking and lung cancer is incomplete and additional information is needed for a proper evaluation.

Mortality from lung cancer is much higher in males than in females and the difference has been increasing In the United States in 1930, the ratio of the lung cancer mortality rate among white males to the rate among white females was 1 7. By 1950, this ratio had increased to 4.6. Although accurate information on the smoking habits of the American public is not available, it is generally believed that smoking among women was rather rare at the beginning of the century and did not become a common practice until some twenty years ago Since lung cancer is believed to have a long latent period, even if smoking is a cause of lung cancer, the effect of this change in the smoking habits of women may not become apparent in mortality statistics for another decade or two Since it has been shown that smoking is more common among women with lung cancer than among comparable controls (Doll and Hill [10]), a marked acceleration in the increase in the lung cancer mortality rate for women during the next one or two decades would lend additional support to the hypothesis that smoking is a cause of lung cancer.

The population studies described above will also provide additional evidence If lung cancer should consistently occur more frequently in smokers than in non-smokers, in various sub-groups of the populations studied, the case for a cause-effect relationship will be greatly strengthened. To refute effectively the hypothesis that smoking is a cause of lung cancer would then require a reasonable explanation, other than causation, for the consistently observed association between smoking and lung cancer.

Laboratory investigations are under way to determine whether a carcinogenic substance can be identified in tobacco and whether lung cancer can be induced in animals with tobacco smoke or condensates of tobacco smoke Efforts to induce lung cancer in animals by inhalation have been unsuccessful to date However, Wynder, Graham, and Croninger [29] recently reported that 44 per cent of 81 mice painted with tobacco tar developed true carcinoma of the skin The production of any form of cancer in laboratory animals with tobacco products may lead to identification of the carcinogenic fraction

Successful production of lung cancer in animals with tobacco would tend to support the hypothesis that smoking causes lung cancer in humans. However, success in producing cancer in an animal does not prove that the same substance is carcinogenic to man Conversely, failure to produce cancer in an animal with a specified carcinogen does not prove that the substance is not carcinogenic to man It is of interest,

that although it is generally accepted that the chromate industry involves a marked lung cancer hazard, Levin [18] reports that "application or injection into animals of various materials to which workers are exposed has thus far failed to produce malignant lung tumors."

SUMMARY

The available evidence on the relationship between smoking and lung cancer is of four kinds:

1. The observed concomitant increase in recorded mortality from lung cancer and consumption of cigarettes;
2. Fourteen case history studies indicating a smaller percentage of non-smokers and a higher percentage of heavy smokers among lung cancer patients than among comparable controls;
3. Preliminary results of two population studies indicating a higher incidence of lung cancer in smokers than in non-smokers, and a still higher incidence in heavy smokers; and
4. The successful production, by at least one team of investigators, of skin cancer in animals with condensates of tobacco smoke.

There is disagreement whether the evidence at hand warrants a conclusion that smoking and lung cancer are causally related. As additional evidence is gathered from observation of human populations and from experimentation with animals, conclusions will be reached which should achieve general acceptance.

REFERENCES

[1] Berkson, Joseph, "Limitations of the application of fourfold table analysis to hospital data," *Biometrics Bulletin*, 2 (1946), 47–53.

[2] Breslow, Lester, Hoaglin, LeMar, Rasmussen, Gladys, and Abrams, Herbert K., "Occupations and cigarette smoking as factors in lung cancer," *American Journal of Public Health*, 44 (1954), 171–81.

[3] "Cancer of the lung (Endemiology), a symposium," *Council for International Organizations of Medical Sciences*, Paris (1953), 18.

[4] Clemmesen, Johannes, Nielsen, Arne, and Jensen, Emmerik, "Mortality and incidence of cancer of the lung in Denmark and some other countries," *Acta Union Internationale contre le Cancer*, IX (1953), 605–36.

[5] Cochran, William G., Mosteller, Frederick, and Tukey, John W., "Principles of sampling," *Journal of the American Statistical Association*, 49 (1954), 13–35.

[6] Cornfield, Jerome, "A method of estimating comparative rates from clinical data. Applications to cancer of the lung, breast and cervix," *Journal of the National Cancer Institute*, 11 (1951), 1269–75.

[7] Cutler, Sidney J., and Loveland, Donald B., "The risk of developing lung cancer and its relationship to smoking," *Journal of the National Cancer Institute*, 15 (1954), 201–11

[8] Doll, Richard, "Bronchial carcinoma; incidence and aetiology," *British Medical Journal*, 2 (1953), 521–27, 585–90

[9] Doll, Richard, and Hill, A Bradford, "The mortality of doctors in relation to their smoking habits, preliminary report," *British Medical Journal*, 1 (1954), 1451-5
[10] Doll, Richard, and Hill, A Bradford, "A study of the aetiology of carcinoma of the lung," *British Medical Journal*, 2 (1952), 1271-86
[11] Dorn, Harold F., "The increase in cancer of the lung," *Industrial Medicine and Surgery*, 23 (1954), 253-7
[12] Dunn, Halbert L, "Lung cancer in the twentieth century," to be published in *Journal of the International College of Surgeons.*
[13] Hammond, E Cuyler, and Horn, Daniel, "The relationship between human smoking habits and death rates," *Journal of the American Medical Association*, 155 (1954), 1316-28
[14] Heller, J R, "Lung cancer," *Journal of Chronic Diseases*, 1 (1955), 96-9
[15] Hueper, W C, "Environmental lung cancer," *Proceedings of First Conference of the National Cancer Institute of Canada*, Academy Press, in press
[16] Korteweg, R, "The age curve in lung cancer," *The British Journal of Cancer*, V (1951), 21-70
[17] Koulumies, Marja, "Smoking and pulmonary cancer," *Acta Radiologica*, 39 (1953), 255-60
[18] Levin, Morton L, "Etiology of lung cancer, present status," *New York State Journal of Medicine*, 54 (1954), 769-77
[19] McConnel, R B, Gordon, K C T, and Jones, T, "Occupational and personal factors in the aetiology of carcinoma of the lung," *Lancet* 2 (1952), 651-6
[20] Mills, Clarence A, and Porter, Marjorie Mills, "Tobacco smoking habits and cancer of mouth and respiratory system," *Cancer Research*, 10 (1950), 539-42
[21] Muller, F. H, "Tabakmissbrauch und Lungencarcinom," *Zeitschrift für Krebsforschung*, 49 (1939), 57-85
[22] Sadowsky, Doris A, Gilliam, Alexander G, and Cornfield, Jerome, "The statistical association between smoking and carcinoma of the lung," *Journal of the National Cancer Institute*, 13 (1953), 1237-58
[23] Schairer, E and Schoniger, E, "Lungenkrebs und Tabakverbrauch," *Zeitschrift fur Krebsforschung*, 54 (1943), 261
[24] Schrek, Robert, Baker, Lyle A, Ballard, George P, Dolgoff, Sidney, "Tobacco smoking as an etiologic factor in disease, I Cancer," *Cancer Research*, 10 (1950), 49-58
[25] Wassink, W F, "Ontstaansvoorwaarden voor Longkanker," *Nederlandsch Tijdschrift voor Geneeskunde*, 92 (1948), 3732
[26] Watson, William L, and Conte, Alexander J, "Smoking and lung cancer," *Cancer* 7 (1954), 245-9
[27] Wynder, Ernest L, and Cornfield, Jerome, "Cancer of the lung in physicians," *New England Journal of Medicine*, 248 (1953), 441-4
[28] Wynder, Ernest L, and Graham, Evarts A, "Tobacco smoking as possible etiologic factor in bronchiogenic carcinoma," *Journal of the American Medical Association*, 143 (1950), 329-36.
[29] Wynder, Ernest L, Graham, Evarts A, and Croninger, Adelle B, "Experimental production of cancer with cigarette tar," *Cancer Research*, 13 (1953), 855-864

NONWHITE POPULATION INCREASES IN METROPOLITAN AREAS*

PAUL F COE
Federal Housing Administration

THIS analysis is devoted primarily to nonwhite population changes which have occurred in standard metropolitan areas (SMA's) from 1940 to 1950. However, a helpful backdrop against which to evaluate the recent SMA nonwhite changes is the long-term growth pattern of nonwhites in the United States total. The need for this orientation is the greater, since the recent trend runs noticeably counter to that which obtained for over a century

NONWHITE POPULATION TRENDS IN THE UNITED STATES

The nonwhite[1] population in the Nation increased by 17 per cent over the last decade to almost 16,000,000 or over 10 per cent of the total of all races, as is shown in Table 1 Thus, for the second successive decade the relative increase in nonwhites exceeded that of whites, and at a conspicuously accelerated rate This marks a departure from the long-standing trend toward a declining proportion of nonwhites which began in 1810 and continued without interruption up to 1930 These recent relative nonwhite population gains are attributable, on the one hand, to a higher rate of natural increase in the nonwhite than in the white population [2] Both the birth and the death rates, the two components of natural increase, have improved materially in recent decades The much greater relative improvement in the nonwhite death rate over the decades, however, accounts for an important part of the recent relative increase in nonwhites

On the other hand, the reduction in white immigration explains a significant part of the recent relative increase in the nonwhite population. Although during the past two decades immigration has been a relatively unimportant factor in our national growth, amounting to about 5 per cent of our total population increase, from 1830 to 1930 it accounted for about one-third of the total population increase and exceeded one-half from 1900 to 1910 [3] Certainly after 1860 the bulk of

* Acknowledgement is made of Austin R Speake's valuable contribution in assisting in the preparation of many of the statistics which underlie this analysis.

[1] The term "nonwhite" consists of Negroes, Indians, Japanese, Chinese, Filipinos, Koreans, Asiatic Indians, Polynesians, and other Asiatics Persons of Mexican birth or ancestry who are not definitely Indian or of other nonwhite race were classified as white

[2] U S Bureau of the Census, *Statistical Abstract of the United States*, 1953, Tables 56 and 67

[3] U S Immigration and Naturalization Service, *1952 Annual Report*, Table 1, and Table 1 of this paper

TABLE 1

TREND OF NONWHITE POPULATION, 1790–1950, AND TREND OF NONWHITE URBAN AND RURAL POPULATION, 1930–1950

Year and residence	Absolute number			Nonwhite as a % of all races	Per cent incre[illegible] over dec[illegible]		
	Nonwhite	White	All races		Non-white	White	[illegible]
U. S total							
1950	15,755,333	134,942,028	150,697,361	10 5	17 1	14 1	14 5
1940	13,454,405	118,214,870	131,669,275	10 2	7 7	7 2	7 2
1930	12,488,306	110,286,740	122,775,046	10 2	14 7	16 3	16 1
1920	10,889,705	94,820,915	105,710,620	10 3	6 3	16.0	14.9
1910	10,240,309	81,731,957	91,972,266	11 1	11 5	22 3	21.0
1900	9,185,379	66,809,196	75,994,575	12 1	17 1	21 2	20 7
1890	7,846,456	55,101,258	62,947,714	12 5	16 2	27 0	25 5
1880	6,752,813	43,402,970	50,155,783	13 5	23 2	26 4	26 0
1870*	5,481,157	34,337,292	39,818,449	13 8	21 2	27 5	26 6
1860	4,520,784	26,922,537	31,443,321	14 4	24 2	37 7	35 6
1850	3,638,808	19,553,068	23,191,876	15 7	26 6	37 7	35 9
1840	2,873,648	14,195,805	17,069,453	16 8	23 4	34 7	32 7
1830	2,328,642	10,537,378	12,866,020	18 1	31 4	33 9	33 5
1820	1,771,656	7,866,797	9,638,453	18 4	28 6	34.2	33 1
1810	1,377,808	5,862,073	7,239,881	19 0	37 5	36 1	36 4
1800	1,002,037	4,306,446	5,308,483	18 9	32 3	35 8	35.1
1790	757,208	3,172,006	3,929,214	19 3	—	—	—
Total nonfarm (Urban and rural nonfarm)							
1950	12,422,237	115,226,774	127,649,011	9 7	42 8	24 2	25 8
1950†	12,418,743	115,202,079	127,620,822	9 7	42 7	24 2	25 8
1940	8,701,679	92,751,408	101,453,087	8 6	15 1	9 0	9 5
1930	7,557,038	85,060,495	92,617,533	8 2	—	—	—
Urban							
1950	9,711,251	86,756,435	96,467,686	10 1	50 5	27 6	29 6
1950†	9,259,600	79,667,864	88,927,464	10 4	43 5	17 2	19 5
1940	6,450,879	67,972,823	74,423,702	8 7	19 6	6 9	7 9
1930	5,394,790	63,560,033	68,954,823	7 8	—	—	—
Rural nonfarm							
1950	2,710,986	28,470,339	31,181,325	8 7	20 4	14 9	15 4
1950†	3,159,143	35,534,215	38,693,358	8 2	40 4	43 4	43 2
1940	2,250,800	24,778,585	27,029,385	8 3	4 1	15 2	14 2
1930	2,162,248	21,500,462	23,662,710	9 1	—	—	—
Rural farm							
1950	3,333,096	19,715,254	23,048,350	14.5	−29.9	−22.6	−23.7
1950†	3,336,590	19,739,949	23,076,539	14 5	−29 8	−22 5	−23 6
1940	4,752,726	25,463,462	30,216,188	15 7	− 3.6	.9	2
1930	4,931,268	25,226,245	30,157,513	16.4	—	—	—

* Adjusted for underenumeration as estimated in the 1930 *Census of Population*, Vol II, Chapter 2, Table 4.

† Old urban definition.

Source U S Bureau of the Census, 1950 *Census of Population* Report P-A1, Table 2 P-B1, Table 34; and *Historical Statistics of the United States*, 1789–1945, Series B 13–23

these immigrants were white and therefore augmented the white increase as compared with the nonwhite.[4] This tended to obscure the strong underlying element of natural increase present among nonwhites for many decades Indeed, had the immigration aspect of the popula-[illegible]crease been eliminated from the white and nonwhite trends of [illegible] century and a half, the nonwhite relative increase probably [illegible]ave fairly closely approximated that of the white during the [illegible]eriod.

For some of the past decades, the nonwhite population counts apparently reflect a sizeable absolute amount of underenumeration.[5] Although the 1870 nonwhite figures shown in Table 1 have been adjusted upward to reflect Census Bureau estimates of that underenumeration, the adjusted fluctuations for that date are still inexplicably erratic, as are those for 1890 and 1920, which also appear too low Those fluctuations are not substantial enough, however, to affect materially the trends and general conclusions pointed out. Moreover, the tendency to undercount nonwhites, especially children under five years of age, is a continuing bias which therefore has little influence on the relative rate of change among nonwhites over the decades [6]

SHIFTS TO NONFARM AREAS

Not all types of areas shared equally in the general upsurge of nonwhite population In fact, in the redistribution of the nonwhite population from 1940 to 1950, those living in rural farm areas declined by almost 30 per cent or by 1,400,000, while the white population in those areas declined by only 22 per cent or by about 5,700,000,[7] as shown in Table 1. Nonwhites living in urban areas (using old urban definition for comparability) and in rural nonfarm areas, on the other hand, increased sharply and at somewhat similar rates, 44 and 40 per cent. However, in absolute terms, nonwhites in urban areas gained about

---

[4] Gunnar Myrdal, *An American Dilemma* (New York) Harper (1944), 119, Henry C Carey, *The Slave Trade* (1853), p. 18, U S. Bureau of the Census, *Historical Statistics of the United States*, 1789–1945, Tables B 304–330, and Helen F Eckerson and Gertrude D Krichefsky, "A Quarter Century of Quota Restriction," *Monthly Review of Immigration and Naturalization Service*, January 1950, p. 91.

[5] U. S Bureau of the Census, 1930 *Census of Population*, Volume II, General Report, p. 26 and *Negro Population in the United States*, 1790–1915, Chapter II.

[6] U. S. Bureau of the Census, 1950 *Census of Population*, Volume I, *Number of Inhabitants*, p. xiii. See also footnote 5.

[7] According to the 1950 *Census of Population*, Volume II, Part 1, pp 33–35, the 1950 definition of farm population differs from that of 1940 and 1930 in that in 1950 persons living on what might have been considered farm land were classified as nonfarm if they paid cash rent for their homes and yards, as also were persons in institutions, summer camps, motels, and tourist camps. There is evidence that the farm population in 1950 would have been about 9 per cent larger had the 1940 classification been used. By appropriately augmenting the 1950 farm figures in order to arrive at an estimate of the number of nonwhites and whites living in rural farm areas on a basis comparable with the 1940 definition, the nonwhite decline would be 24 per cent, or 1,100,000, the comparable percentage decline in white persons living on farms is estimated to be almost 16 per cent, or about 4,000,000.

2,800,000 compared with 900,000 for the rural nonfarm classification. The relative increase in the white population in the rural nonfarm areas was at about the same rate as was that of the nonwhites, but in urban places, the white increase was considerably less than half as great relatively, 17 per cent, as was the nonwhite

As a result of these population shifts together with the differing rates of natural increase, over 12,400,000 nonwhites lived in nonfarm areas in 1950, which amounts to almost 79 per cent of the total nonwhites in the United States as compared with 65 per cent in 1940 and 61 per cent in 1930 The white population in nonfarm areas, still relatively greater than nonwhite, increased to over 85 per cent of the United States total count of whites in 1950, as compared with 78 per cent in 1940 and 77 per cent in 1930 Percentage-wise, therefore, the nonwhites in nonfarm areas have been gaining rapidly over the past two decades as compared with the white population Using the new urban definition,[8] however, nonwhites have increased much faster relatively than they have in the rural nonfarm areas These trends clearly indicate the sharp movement of the nonwhite population away from farms and especially to urban areas which are the industrial centers of the Nation

NONWHITE SHIFTS TO SMA'S

Turning now to population changes in SMA's, which is the main theme of this paper, the rapid surge of nonwhite population into SMA's during the 1940's is one of the most significant and dramatic population trends revealed by the last census It is appropriate, therefore, to examine the attributes of the SMA, in order to understand better the type of locality to which the nonwhites are attracted so strongly As a class, SMA's possess the highest degree of urbanization of all types of areas. They are the Nation's big cities and their environs Formally, the U. S Bureau of the Budget has defined an SMA as a county or group of contiguous, socially and economically integrated counties, which contains at least one city of 50,000 or more inhabitants, except in New England where somewhat different criteria are employed [9]

[8] According to the 1950 *Census of Population*, Volume II, Part 1, Chapter B, p VI, the 1950 definition of urban population was expanded to include as urban, persons living (a) in the densely settled urban fringe, both incorporated and unincorporated, surrounding cities of 50,000 population and (b) in other unincorporated places of 2,500 or more This change in definition resulted in an increase in the urban and a roughly compensating decrease in the rural nonfarm population count of nonwhites of about 450,000 and of whites of almost 7,100,000 in 1950 Thus, the non-white totals were affected less, both relatively and absolutely than were the white by the change in definition

[9] In addition to the county, or counties, containing a central city, or cities, of 50,000 population,

In order to highlight the trend of nonwhites to SMA's, a table showing pertinent statistics of nonwhite population changes in each of the 168 SMA's has been prepared by the Market Analysis Section, Division of Research and Statistics, FHA. The summary data which that table presents for the United States, as well as the separate SMA data for the Washington, D C, area, are reproduced in Table 2. The Washington, D C, data in this summary table serve to illustrate the type of statistics contained in a detailed table embracing each of the 168 SMA's in the continental United States in 1950, which table is available on request from the FHA [10] Because of space limitations, it is not feasible to publish here the statistics for each of the SMA's Moreover, while the detailed table includes data for 1930, this analysis is focused largely on the changes from 1940 to 1950

Of the 15,755,000 nonwhite[11] persons living in the continental United States in 1950, over 8,250,000 lived in SMA's—an SMA increase of 2,534,000 from 1940 An indication of the magnitude of that SMA nonwhite increase can be had by noting that it exceeded the 1950 all race population count of Philadelphia by almost one-half a million. Moreover, it also substantially exceeded the aggregate 1950 count of all persons living in Montana, Idaho, Wyoming, Utah, and Nevada.

Nonwhites in SMA's increased over twice as fast relatively as did the white population, 44 per cent compared with 20 per cent Moreover, nonwhites in SMA's increased over 2½ times as fast percentage-wise as they did in all types of areas in the United States As a result of these changes, approximately 1 of every 10 persons living in SMA's in 1950 was nonwhite From 1930 to 1940 also the nonwhites increased relatively much faster in SMA's than in the country at large, 16 per cent compared with 8 per cent, respectively.

It is estimated that migration accounted for almost two-thirds of

---

contiguous counties are included in an SMA if according to specified criteria they are essentially metropolitan in character and are socially and economically integrated with the central city The criteria of metropolitan nature relate primarily to the character of the county as a place of work or as a home for concentrations of nonagricultural workers and their dependents Since in New England the towns and cities are more important units administratively than is the county, they are the units in terms of which those SMA's are delineated In New England, the criterion of a minimum population density of 150 persons per square mile also applies in most instances The definition of an SMA is given more fully in the U S Bureau of the Census, 1950 *Census of Population, Number of Inhabitants, U S Summary*, Report P-A1, p XXXI, for a precise delineation of each SMA, see Tables 26 and 27 of the U S Bureau of the Census, Report P-A1

[10] Single copies of the complete table containing data for each of the 168 standard metropolitan areas may be obtained without charge, while the supply lasts, by writing to the Division of Research and Statistics, Federal Housing Administration, Washington, D C

[11] Of the nonwhite total, 15,042,286, or 95 5 per cent, were Negroes The greater number of the remaining 713,047 nonwhites were American Indians, with Japanese and Chinese next most numerous

TABLE 2

TREND OF NONWHITE POPULATION IN THE 168 STANDARD METROPOLITAN AREAS, IN CENTRAL CITIES, OUTSIDE CENTRAL CITIES, AND IN THE WASHINGTON SMA, 1930–1950

Color and year		U S summary data					Washington, D C. SMA (Similar data available for each of the 168 SMA's)		
		Continental U. S total population	Total 168 SMA's: Total	Total 168 SMA's: In central cities	Total 168 SMA's: Outside central cities	Total outside SMA's	Total	In central cities	Outside central cities
All races:	1950	150,697,361	84,500,680	49,412,792	35,087,888	66,196,681	1,464,089	802,178	661,911
	1940	131,669,275	69,276,481	43,391,718	25,884,763	62,392,794	967,985	663,091	304,894
	1930	122,775,046	63,831,815	41,184,908	22,646,907	58,943,231	672,198	486,869	185,329
Population									
Nonwhite	1950	15,755,333	8,250,210	6,411,158	1,839,052	7,505,123	342,159	284,313	57,846
	1940	13,454,405	5,716,537	4,323,644	1,392,893	7,737,868	230,827	188,765	42,062
	1930	12,488,306	4,913,703	3,624,504	1,289,199	7,574,603	168,294	132,888	35,406
Population increase									
White	1940–50	16,727,158	12,690,526	3,933,560	8,756,966	4,036,632	384,772	43,539	341,233
	1930–40	7,928,130	4,641,832	1,507,670	3,134,162	3,286,298	233,254	120,345	112,909
Nonwhite	1940–50	2,300,928	2,533,673	2,087,514	446,159	−232,745	111,332	95,548	15,784
	1930–40	966,099	802,834	699,140	103,694	163,265	62,533	55,877	6,656
Percent increase									
All races·	1940–50	14.5	22.0	13.9	35.6	6.1	51.3	21.0	117.1
	1930–40	7.2	8.5	5.4	14.3	5.9	44.0	36.2	64.5
White·	1940–50	14.1	20.0	10.1	35.8	7.4	52.2	9.2	129.8
	1930–40	7.2	7.9	4.0	14.7	6.4	46.3	34.0	75.3
Nonwhite	1940–50	17.1	44.3	48.3	32.0	−3.0	48.2	50.6	37.5
	1930–40	7.7	16.3	19.3	8.0	2.2	37.2	42.0	18.8
Nonwhite population as a % of all races	1950	10.5	9.8	13.0	5.2	11.3	23.4	35.4	8.7
	1940	10.2	8.3	10.0	5.4	12.4	23.8	28.5	13.8
	1930	10.2	7.7	8.8	5.7	12.9	25.0	27.3	19.1
Rank in 1950 population									
Nonwhite		—	—	—	—	—	5	—	—
All races		—	—	—	—	—	11	—	—

Source The data in this and in the following tables were derived from the FHA table entitled, "Nonwhite Population Trends in Standard Metropolitan Areas, in Central Cities, and Outside Central Cities, 1930–1950." The primary sources of these data are, U S Bureau of the Census, 1950 *Census of Population*, Report P-A, Numbers 2–50, Table 4; Report P-B, Numbers 2–50, Table 34, Report P-A1, Tables 26 and 27; Report P-B1, Table 39, and Series P-C14; 1940 *Census of Population, Characteristics of the Population*, Vol II, Tables 21, 28, 29, 30, 31, and 36

the nonwhite population increase in SMA's[11] and in their central cities, and for almost half of the nonwhite increase in that part of the SMA's lying outside the central cities (referred to hereafter as suburbs, even though some of this area is open country). This may be inferred from the fact that the nonwhite natural increase in the United States total from 1940 to 1950 amounted to 17.1 per cent, whereas the nonwhite population increase was 44.3 per cent in SMA's, 48.3 per cent in their central cities, and 32 per cent in the suburbs. A rough estimate of the per cent of nonwhite in-migration can be computed indirectly by algebraically subtracting from each of the three foregoing percentage increases, the 17.1 per cent rate of nonwhite natural increase in the U. S. total and by then dividing that percentage difference by the total per cent increase in nonwhites in each of the three classifications. Accordingly, in-migration from 1940 to 1950 is estimated to be about 61 per cent of the nonwhite population increase in the 168 SMA's, about 65 per cent of all central cities, and about 47 per cent in the suburbs of all SMA's. It is obvious that the source of the migration was the nonwhite population living outside SMA's, inasmuch as that segment of the nonwhite population of the United States actually declined by 233,000, or 3 per cent, in contrast to the 17.1 per cent natural increase noted for the total nonwhite population. The nonmetropolitan segment remains an important potential source of further nonwhite migration, inasmuch as 7,505,000 nonwhites, or 48 per cent of the United States nonwhite total, still lived outside SMA's in 1950

In the Washington, D. C SMA, the locality for which exhibit data are shown in Table 2, nonwhite in-migration percentages[12] somewhat exceeded those noted in the foregoing paragraph for the total of all 168 SMA's in the U. S Thus, it is estimated that about 65 per cent of the total nonwhite decennial increase in the Washington SMA sprang from in-migration, as did about 66 per cent in the city proper, and about

---

[12] Nonwhite birth and death statistics are reported annually for most SMA counties and central cities in the source given in footnote 13. By assembling the appropriate county natural increase figures for the years 1940 through 1949 and by then subtracting them from the population increase over the decade (given in the table referred to in footnote 10) the absolute and relative quantity of nonwhite in-migration can be computed fairly accurately for any SMA. In contrast, the crude but very convenient method of estimating nonwhite and white in-migration described in this paper yields only approximate results for the reason that the rate of net natural increase inside SMA's is less than that of the population living outside SMA's Moreover, the rate of net natural increase varies noticeably among the 168 SMA's. Consequently, the validity of this indirect method of estimating in-migration may vary substantially among the SMA's. It is likely that this method will yield useful and fairly accurate estimates of in-migration for SMA's which experienced a high rate of population increase from 1940 to 1950. Conversely, for SMA's which fell substantially below the average rate of population increase of the Nation, this device may very likely produce untrustworthy estimates of migration. Note that the estimate of nonwhite in-migration computed by this method for the Washington, D C SMA closely approximated that developed from natural increase statistics.

54 per cent in the metropolitan fringe. Corresponding figures for the white population indicate that about 73 per cent of the total white increase in this SMA in-migrated, as did about 89 per cent in the SMA fringe, whereas about 53 per cent of the net white natural increase in the city proper out-migrated—in contrast to the 66 per cent in-migration noted for nonwhites in the city In some of the other specific SMA's the importance of migration varied widely from that of the Washington, D. C. SMA, and from the aggregate total of all 168 SMA's Decennial natural increase statistics by color, required in the exact derivation of in-migration estimates, have not been developed here for any SMA other than Washington, D C. However, a rough approximation of the percentage increase ascribable to in-migration can be computed inferentially for any other SMA by using the method discussed in the previous paragraph (Individual SMA percentage increases are available from the source indicated in footnote 10) Incidentally, in the Washington, D. C. SMA the actual rate of natural increase in nonwhites over the decade was computed to be 15 7 per cent[13] before upward adjustment for underregistration, quite close to the 17 1 per cent natural increase noted for the U. S total.

INCREASES INSIDE CENTRAL CITIES

Nonwhite persons have been gravitating rapidly to the large centers of population, and more particularly to the central cities of those areas. Thus, the nonwhites inside the central cities of the 168 SMA's increased by 2,088,000, compared with 446,000 in the SMA suburbs The movement further concentrated the nonwhites in the congested areas of the cities, so that in 1950 there were 6,411,000 nonwhites inside central cities and only 1,839,000 in the suburbs. While nonwhites inside all central cities increased by 48 per cent, the comparable white population increased by only 10 per cent As the SMA's grew, the population overflowed the central city boundaries and into the suburbs. Thus, in the suburbs, the nonwhites increased by 32 per cent and the whites by 36 per cent, as is shown in Table 2.

As a result of this redistribution, over half of the nonwhite population in the United States, 52 per cent, resided in SMA's in 1950, compared with 42 per cent in 1940 Over the 10-year period, the proportion of the total white population living in SMA's grew to 57 per cent in 1950 from 54 per cent in 1940

These recent nonwhite shifts were under way a decade earlier, for

[13] Natural increase figures employed in these computations were compiled from the U. S. National Office of Vital Statistics, *Vital Statistics of the U. S., Part II, Place of Residence*, annual volumes, 1940 through 1949.

from 1930 to 1940 the nonwhites inside central cities increased by 19 per cent whereas their increase in the SMA suburbs was only 8 per cent.

The further concentration of nonwhites in central cities is not unlike the experience of earlier large-scale migrations, especially of white immigrants during the second half of the 19th century. Much of the housing occupied by in-migrant nonwhites was formerly occupied by the white population living in neighborhoods adjacent to established nonwhite neighborhoods. This is partly because of an apparent general but not universal preference on the part of nonwhites for close-in locations with their available church, social, and entertainment facilities, rent differentials, proximity to sources of employment, and transportation advantages; it is also partly because of a lack of available new construction for nonwhites in the outlying areas. In turn, the white households which had occupied housing in transition to nonwhite occupancy relocate to a significant degree in the outskirts of the city proper or in the suburbs, as is demonstrated by figures given in subsequent paragraphs

EXPANDING CENTRAL CITY BOUNDARIES

It is likely that some of these decennial changes as between inside central cities and their suburbs were apparent rather than real, being partly attributable to the expansion of the boundaries of many central cities with a compensating contraction of that part of the SMA lying outside the central cities. This follows from the fact that the land area encompassed by these 193 central cities (contained in the 168 SMA's) increased from 5,720 square miles in 1940 to 6,573 square miles in 1950, or by 15 per cent over the decade [14] In 11 of these 193 central cities, most of which were in the West or in the South, there were an-

[14] U. S Bureau of the Census, 1940 *Census of Population*, Volume I, *Number of Inhabitants*, Table 17, and *Areas of the United States*, Table 4; 1950 *Census of Population*, *Land Area and Population of Incorporated Places of 2,500 or More*, Series GEO No 5. A compilation from these sources yields the following distribution of change in number of square miles from 1940 to 1950—

Change in number of square miles	*Number of central cities*
All cities	*193*
Decrease	6
No change	69
Increase	*118*
0 1 to 0 9	30
1 0 to 1 9	22
2 0 to 4 9	28
5.0 to 9.9	13
10.0 to 19 9	14
20 0 to 49 9	8
50.0 to 89.9	3

nexations ranging from 20 to 87 square miles. However, 121 central cities underwent no change or increased by less than 2 square miles. Generally, the suburban areas contained relatively fewer nonwhites than did the central cities to which these areas were annexed. Inasmuch as in 1940 the population density of the central cities was 15 times as great as that of the suburbs, the 15 per cent increase in area of central cities undoubtedly accounts for a much smaller, but unknown, per cent increase in the population of these central cities.

It should be emphasized, of course, that the suburban areas into which the corporate limits of the central cities overflow, tend to exhibit the same urban characteristics as do the adjacent areas of the central cities themselves Therefore, usually the formal act of incorporation of a suburban segment simply recognizes legally and administratively at irregular intervals that condition of urban development which proceeds at a more regular pace and according to more orderly socio-economic laws. Although annexations enlarged the area of 118 of the 193 central cities, the over-all boundaries of the SMA's themselves were held constant from 1930 to 1950 in the statistical comparisons presented in this analysis.

CENTRAL CITIES LOSING WHITE AND GAINING NONWHITE POPULATION

In 22 central cities of SMA's having an all race population of close to 9,000,000 in 1950, the nonwhite population increased by about 446, 000 or 58 per cent, while the white population decreased by about 142,000, or almost 2 per cent from 1940 to 1950, as is shown in Table 3. These increase-decrease relationships, however, varied widely.

In some cities these shifts were quite substantial. The city of Chicago, for example, lost 3,000 white persons but gained 227,000 nonwhite. Comparable figures for St Louis are 4,000 white lost, compared with 45,000 nonwhite gained, Cleveland, 28,000 white lost, 65,000 nonwhite gained, Pittsburgh, 15,000 white lost, 21,000 nonwhite gained; Newark, 20,000 white lost, 29,000 nonwhite gained; and Buffalo, 15,000 white lost, 19,000 nonwhite gained.

In the suburbs surrounding each of these 6 cities the numerical increase in white population was at least 8 times as great as the nonwhite increase.

In six of these 22 central cities, the white population decreased by over 5.0 per cent, with a decline of 7.3 per cent in Atlantic City. In all but three of the 22 central cities the nonwhite population increased much faster relatively than the white lost. In only 7 of these 22 cities was the all race population less than 100,000.

TABLE 3

THE 22 SMA CENTRAL CITIES IN WHICH NONWHITES INCREASED AND WHITES DECREASED, RANKED BY WHITE PER CENT DECREASE, 1940–1950

Central cities	All races 1950	White decrease 1940–50	Nonwhite increase 1940–50	Per cent change 1940–50		Nonwhite as a % of all races
				White	Nonwhite	
Atlantic City, N. J.	61,657	−3,552	1,115	−7.3	7.1	27.3
Hazleton, Pa.	35,491	−2,536	18	−6.7	900.0	.1
Johnstown, Pa	63,232	−4,079	643	−6.3	40.8	3.5
Steubenville, Ohio	35,872	−2,011	232	−5.8	8.2	8.5
Newark, N. J.	438,776	−20,385	29,401	−5.3	63.6	17.2
Wilmington, Del.	110,356	−5,096	2,948	−5.2	20.6	15.7
Lawrence, Mass	80,536	−3,898	111	−4.6	74.0	.3
Lowell, Mass.	97,249	−4,212	72	−4.2	52.6	.2
Youngstown, Ohio	168,330	−6,273	6,883	−4.1	46.9	12.3
Wheeling, W. Va.	58,891	−2,303	95	−3.9	5.0	3.4
Cleveland, Ohio	914,808	−28,153	64,625	−3.5	76.1	16.3
Jersey City, N. J.	299,017	−9,547	7,391	−3.3	54.4	7.0
Providence, R. I.	248,674	−6,911	2,081	−2.8	31.5	3.5
Buffalo, N. Y.	580,132	−15,186	19,417	−2.7	106.2	6.5
Pittsburgh, Pa.	676,806	−15,411	20,558	−2.5	32.9	12.3
Reading, Pa	109,320	−2,262	1,014	−2.1	52.8	2.7
Trenton, N. J.	128,009	−1,880	5,192	−1.6	55.6	11.4
St. Louis, Mo	856,796	−4,446	45,194	− .6	41.4	18.0
Nashville, Tenn	174,307	−491	7,396	− .4	15.6	31.4
Chicago, Ill.	3,620,962	−3,039	227,193	− .1	80.5	14.1
Utica, N. Y.	101,531	−128	1,141	− .1	215.7	1.6
Chattanooga, Tenn	131,041	−22	2,900	*	8.0	30.0
Total	8,991,793	−141,821	445,620	−1.8	57.8	13.5

* Less than 0.05 per cent.

Nonwhites in these central cities were present in about the same proportion, 13.5 per cent, as they were in the central cities of all 168 SMA's, 13.0 per cent. Yet in Nashville, Chattanooga, and Atlantic City they were more than twice as numerous relatively. The majority of these 22 cities are located in the Northeastern Region of the Nation. Only 3 are in the South.

REGIONAL SHIFTS

Although marked shifts in the nonwhite population occurred among the four regions of the United States during the decade, the South still had, in 1950, by far the largest number of nonwhites living in SMA's, 3,577,000. Moreover, nonwhites were over twice as numerous propor-

tionately in southern SMA's as in the other regions, as is shown in Table 4.

The North Central SMA's experienced the greatest absolute increase in nonwhite population, 810,000 since 1940 That increase brought the nonwhites to 2,022,000 in 1950, or over 8 per cent of all SMA population in that region It is important to note, however, that war and post-

TABLE 4

REGIONAL TRENDS IN THE NONWHITE POPULATION IN THE 168 SMA'S, 1940–1950

Year and color	Total 168 SMA's	Population in SMA's of each region			
		North-east	North Central	South	West
All races, 1950	84,500,680	30,891,820	24,491,036	17,200,809	11,917,015
Nonwhite, 1950	8,250,210	1,953,399	2,021,691	3,577,208	697,912
Population increase 1940–1950					
White	12,690,526	2,210,306	2,902,505	3,892,188	3,685,527
Nonwhite	2,533,673	646,298	809,836	686,315	391,224
Per cent increase 1940–1950					
Total	22 0	10 2	17 9	36 3	52 0
White	20 0	8 3	14 8	40 0	48 9
Nonwhite	44 3	49 4	66 8	23 7	127 6
Nonwhite as per cent of total					
1950	9 8	6 3	8 3	20 8	5 9
1940	8 3	4 7	5 8	22 9	3 9
Number of SMA's 1950	168	39	53	58	18

war employment opportunities resulted in very large absolute increases in the number of nonwhites living in the SMA's of each region.

The average rate of nonwhite increase in SMA's was greatest in the West, 128 per cent. There were, however, substantial relative nonwhite SMA increases in the North Central Region, 67 per cent, and in the Northeast, 49 per cent The white population in the SMA's of each of these three regions, in contrast, increased at a much lower rate than did the nonwhite. Yet it is quite significant that in the SMA's of the South the white population increased almost twice as fast relatively as did the nonwhite, 40 per cent compared with 24 per cent.

The observed tendency of the nonwhite population to increase relatively faster than the white was quite general among the SMA's in each region outside the South. It occurred in 34 of the 39 SMA's in the Northeast, in 14 of the 18 SMA's of the West, and in 46 of the 53

SMA's of the North Central Region. Contrary-wise, in 51 of the 58 SMA's in the South, the nonwhite population increased at a slower rate than did the white.

HIGHEST PERCENTAGE NONWHITE

Each of the 10 SMA's with 50,000 or more nonwhites in 1950 and with nonwhites comprising at least one-third of the total population was located in the South, as is shown in Table 5. The percentage of nonwhite to total population, however, declined over the decade in each of these 10 SMA's. The relative increase in number of white persons was much greater than that of nonwhites in each of these SMA's. Nevertheless, it is important that, despite the more rapid relative nonwhite SMA gains in other parts of the country, there was an increase of over 10 per cent in 8 of these 10 southern SMA's In the Baton Rouge and Mobile SMA's, nonwhites increased by over 50 per cent during the decade.

TABLE 5

THE 10 SMA'S WITH MORE THAN 50,000 NONWHITES IN WHICH NONWHITES COMPRISED MORE THAN 33 PER CENT OF THE ALL RACE TOTAL, RANKED BY PER CENT NONWHITE, 1950

Standard metropolitan area	Nonwhite population 1950	Per cent increase 1940–50		Nonwhite population as a % of all races	
		White	Non-white	1950	1940
Jackson, Miss.	63,917	51.0	15 3	45 0	51 7
Montgomery, Ala.	60,616	37 3	5.7	43.6	50.1
Charleston, S C.	68,354	56 9	14.7	41 5	49.2
Savannah, Ga	58,547	42 9	10 6	38.6	44 9
Memphis, Tenn	180,185	48.9	16 0	37.4	43 3
Birmingham, Ala.	208,616	24 8	16 4	37 3	39 0
Columbia, S. C	50,494	47.4	19.2	35 4	40.4
Augusta, Ga	56,113	36.7	3 4	34 6	41.2
Mobile, Ala	77,999	69 6	50 9	33.8	36 4
Baton Rouge, La	52,341	93 3	55.6	33.1	38.0

LARGEST SMA'S

The heavy concentration of nonwhites in a few SMA's is shown by the fact that the 10 SMA's containing the largest number of nonwhites accounted for about half, 49 per cent, of the nonwhites living in all 168

SMA's and for over one-fourth of all nonwhites in the Nation in 1950. More nonwhites lived in the New York SMA alone than in any of 46 States—1 out of every 15 nonwhites in the United States.

The relative increase in nonwhites far exceeded that of whites in these 10 SMA's, 63 compared with 17 per cent, as is shown in Table 6, and also substantially exceeded the nonwhite increase of 44 per cent in all 168 SMA's. Because of this large increase, the nonwhites as a per-cent of the all race total population in these 10 SMA's in 1950 (10 9) exceeded the proportion in all SMA's and in the United States total, 9.8 and 10.5 per cent, respectively. In 1940 the proportion of non-

TABLE 6

THE 22 SMA'S WITH NONWHITE INCREASES OF MORE THAN 20,000, 1940–1950, RANKED BY NUMBER OF NONWHITES

Standard metropolitan area	All races 1940	Nonwhite population—		Per cent increase, 1940–50		Nonwhite as a % of all races, 1950	Nonwhite rank 1950
		1950	Increase, 1940–50	White	Nonwhite		
New York, N Y	12,911,994	1,046,045	377,191	8 0	56 4	8 1	1
Chicago, Ill	5,495,364	605,238	270,373	8 9	80 7	11 0	2
Philadelphia, Pa	3,671,048	483,927	147,084	11 3	43 7	13 2	3
Detroit, Mich.	3,016,197	361,927	189,149	20 4	109 5	12 0	4
Washington, D C	1,464,089	342,159	111,332	52 2	48 2	23 4	5
Los Angeles, Cal	4,367,911	276,330	148,291	46 7	115 8	6 3	6
Baltimore, Md	1,337,373	266,671	71,895	20 5	36 9	19 9	7
St. Louis, Mo.	1,681,281	216,454	65,006	14 4	42 9	12 9	8
San Francisco, Cal	2,240,767	210,547	145,816	45 3	225 3	9 4	9
Birmingham, Ala	558,928	208,616	29,442	24 8	16 4	37 3	10
Total, largest 10	36,744,952	4,017,914	1,555,579	17 2	63 2	10 9	—
New Orleans, La.	685,405	200,523	40,742	23 5	25 5	29 3	11
Memphis, Tenn	482,393	180,185	24,890	48 9	16 0	37 4	12
Atlanta, Ga.	671,797	165,816	22,422	35 0	15 6	24 7	13
Cleveland, Ohio	1,465,511	154,117	65,888	11 2	74 7	10 5	14
Houston, Texas	806,701	150,452	46,310	54 5	44 5	18 7	15
Pittsburgh, Pa.	2,213,236	137,261	24,372	5 4	21 6	6.2	16
Norfolk-Portsmouth, Va.	446,200	122,837	35,481	88 5	40.6	27 5	17
Cincinnati, Ohio	904,402	95,656	26,636	12 6	38 6	10 6	18
Kansas City, Mo.	814,357	88,032	20,330	17 3	30 0	10 8	19
Dallas, Texas	614,799	83,352	21,639	57 8	35 1	13 6	21
Mobile, Ala	231,105	77,999	26,321	69 6	50 9	33 8	23
Buffalo, N. Y.	1,089,230	47,786	23,905	11 4	100 1	4 4	42
Total, next 12	10,425,136	1,504,016	378,936	20 3	33 7	14 4	—
Total, largest 22	47,170,088	5,521,930	1,934,515	17.9	53 9	11.7	—

whites to all races had been smaller in the ranking 10 SMA's than in all SMA's or in the United States as a whole.

The 1,556,000 increase in the number of nonwhites living in the 10 largest SMA's represented 61 per cent of the 2,534,000 nonwhite increase in all SMA's The nonwhite segment of each of these 10 SMA's in 1950 was in fact itself the equivalent of a large city. In the New York SMA, for example, there were 1,046,000 nonwhites in 1950. Only 14 SMA's had a greater all race population than that in 1950. Over 605,000 nonwhites lived in the Chicago SMA in 1950. Even the Birmingham SMA, which ranked 10th in number of nonwhites, had 209,000, a larger number than the all race population in any of 81 SMA's. Of the 10 SMA's with the largest number of nonwhites, only Washington, Baltimore, and Birmingham are located in the South.

By expanding this group to include each SMA in which the nonwhite population increased by 20,000 or more between 1940 and 1950 there are 22 SMA's in all, shown also in Table 6. Of these 12 additional SMA's, seven are located in the South, whereas only 3 of the 10 SMA's having the largest number of nonwhites are in the South Also, of these 12 SMA's, only four exceeded the 44.3 per cent non-white average increase for the 168 SMA's Moreover, in only two of these 12 SMA's, Buffalo and Pittsburgh, did the nonwhites as a per cent of total population fail to equal the 9 8 per cent average for all 168 SMA's. The relative increase in nonwhites for these 12 SMA's was 33 7 per cent, as against 20 3 per cent for the whites. This percentage increase of nonwhites was far below that of the 10 SMA's with the largest number of nonwhites, most of which are located outside the South.

For the 22 SMA's combined, the nonwhite population increased by almost 2,000,000, or 54 per cent, to a little over 5,500,000 which represents 11.7 per cent of all races The comparable white population increase was only 18 per cent. Factors contributing to the substantial movement of the nonwhites to the larger cities include the greater income opportunities in the cities, less need for farm labor with the increased mechanization of the farms and plantations, and personal preference to move to the North and West.

LARGEST RELATIVE INCREASES

Whereas there was only one SMA, Albuquerque, in which the white population doubled from 1940 to 1950, there were 31 SMA's in which the nonwhite population more than doubled The nonwhite increase in these 31 SMA's aggregated over 630,000 or one-fourth of the 2,534,000 nonwhite increase in all 168 SMA's during the decade.

TABLE 7

THE 31 SMA'S IN WHICH NONWHITE POPULATION DOUBLED 1940–1950, RANKED BY HIGHEST PER CENT OF NONWHITE INCREASE

SMA's in which nonwhites doubled	All races 1950	Nonwhite population—		Per cent increase, 1940–50		Nonwhite as a % of all races 1950
		1950	Increase, 1940–50	White	Nonwhite	
SMA's having 10,000 or more nonwhites						
San Francisco, Cal	2,240,767	210,547	145,816	45.3	225.3	9.4
San Diego, Cal	556,808	23,841	14,121	90.6	145.3	4.3
Milwaukee, Wis	871,047	23,241	13,623	12.0	141.6	2.7
Los Angeles, Cal.	4,367,911	276,330	148,291	46.7	115.8	6.3
Portland, Oregon	704,829	15,949	8,484	39.5	113.7	2.3
Denver, Colo	563,832	20,190	10,681	36.5	112.3	3.6
Detroit, Mich	3,016,197	361,927	189,149	20.4	109.5	12.0
Flint, Mich	270,963	14,277	7,451	16.1	109.2	5.3
Fresno, Cal	276,515	19,165	9,754	52.1	103.6	6.9
Buffalo, N Y	1,089,230	47,786	23,905	11.4	100.1	4.4
Total, 10 SMA's	13,958,099	1,013,253	571,275	34.2	129.3	7.3
SMA's having fewer than 10,000 nonwhites						
Racine, Wis	109,585	1,880	1,367	15.2	266.5	1.7
Manchester, N H	88,370	156	105	7.7	205.9	.2
San Bernardino, Cal	281,642	8,641	5,566	72.7	181.0	3.1
Ogden, Utah	83,319	2,038	1,266	45.3	164.0	2.4
Lubbock, Texas	101,048	7,937	4,862	91.2	158.1	7.9
Saginaw, Mich	153,515	9,183	5,570	13.8	154.2	6.0
Grand Rapids, Mich	288,292	7,226	4,324	15.5	149.0	2.5
Erie, Pa	219,388	3,627	2,150	20.3	145.6	1.7
Utica-Rome, N Y	284,262	2,595	1,524	7.5	142.3	.9
Spokane, Wash	221,561	3,052	1,768	33.8	137.7	1.4
South Bend, Ind	205,058	8,831	5,050	24.2	133.6	4.3
Lima, Ohio	88,183	4,403	2,472	17.4	128.0	5.0
Tacoma, Wash	275,876	8,506	4,732	49.9	125.4	3.1
Madison, Wis.	169,357	1,059	588	29.3	124.8	.6
Kalamazoo, Mich	126,707	2,794	1,518	25.4	119.0	2.2
Rochester, N Y	487,632	8,247	4,469	10.3	118.3	1.7
New Britain-Bristol, Conn	146,983	1,549	839	15.4	118.2	1.1
Salt Lake City, Utah	274,895	3,871	2,061	29.2	113.9	1.4
Fort Wayne, Ind.	183,722	5,368	2,795	16.9	108.6	2.9
Springfield-Holyoke, Mass	407,255	7,459	3,826	10.7	105.3	1.8
Peoria, Ill	250,512	6,507	3,317	17.0	104.0	2.6
Total, 21 SMA's	4,447,162	104,929	60,169	22.6	134.4	2.4
Total, 31 SMA's	18,405,261	1,118,182	631,444	31.1	129.7	6.1

Among these 31 SMA's where the nonwhite population more than doubled were 10 with over 10,000 nonwhites in 1950, shown in Table 7. In the San Francisco-Oakland SMA, the nonwhites more than trebled. For these 10 SMA's the nonwhite population increase averaged 129 per cent, compared with 34 per cent for the white Although in most of these 10 largest SMA's the rate of increase for both the nonwhite and the white population was larger than the average rate for all SMA's, in Milwaukee, Flint, and Buffalo the white increase was less than average, while the nonwhites more than doubled On the average, nonwhite population comprised only 7 3 per cent of the total population of these 10 SMA's. All 10 are located outside the South.

Twenty-one of the 31 SMA's in which the number of nonwhites doubled from 1940 to 1950, had fewer than 10,000 nonwhites in 1950. In eleven of the 21 the per cent increase in white population was below the 168 SMA average. For the 21 SMA's, the nonwhite population increase averaged 134 per cent, compared with 23 per cent for the white. The nonwhite population comprised only 2 4 per cent of the total population of the 21 SMA's, compared with 7 3 per cent for the top 10 SMA's.

Of the 31 SMA's in which nonwhites doubled from 1940 to 1950, only Lubbock, Texas, is in the South

NONWHITE PERCENTAGE INCREASE EIGHT TIMES THE WHITE

During the last decade there were 10 SMA's with an all race population of 100,000 or more in which the per cent increase in the nonwhite population was ten or more times that of the white population. Moreover, in 16 SMA's of 100,000 population, the nonwhite percentage increase was eight or more times as great as that of the white, as shown in Table 8 Half of these 16 SMA's had nonwhite percentage increases of over 100 per cent, 2 had increases between 75 and 100 per cent, and only 6 had increases of less than 50 per cent None of these SMA's was located in the South. The average percentage increase for all 16 SMA's was 84 per cent for the nonwhites and 8 per cent for the whites. The nonwhite population comprised 6 6 per cent of that for all races and numbered 733,000 in these 16 SMA's in 1950

VARIATIONS IN RELATIVE GROWTH

Although the increase in nonwhite population in all 168 SMA's averaged 44 per cent from 1940 to 1950, that rate of increase cannot be regarded as typical, as is shown in Table 9 Thus, in only 16 of the 168 SMA's did the increase in nonwhites range between 40 and 50

per cent. In short, variation characterizes the rate of nonwhite growth over the decade, and to a somewhat smaller extent the white growth as well.

The nonwhite changes ranged from a decrease of 37 per cent for the Fall River SMA to an increase of 266 for the Racine SMA. In all, 3 SMA's also showed a decline of over 20 per cent and 3 underwent an increase of over 200 per cent. A total of 9 SMA's recorded a decline in nonwhite population Yet in 61 SMA's the nonwhites increased by 50 per cent or more. The comparable white population increase of over 50 per cent was experienced by only 27 SMA's. Despite the wide range in

TABLE 8

THE 16 SMA'S OF MORE THAN 100,000 POPULATION IN WHICH THE NONWHITE PER CENT INCREASE WAS EIGHT OR MORE TIMES THAT OF THE WHITE, 1940–1950

Standard metropolitan area	Absolute increase		Per cent increase		Nonwhite as a % of all races
	White	Nonwhite	White	Nonwhite	
Albany-Schenectady-Troy, N Y.	44,741	4,106	9 7	82.5	1.8
Buffalo, N Y	106,838	23,905	11 4	100 1	4.4
Chicago, Ill	399,464	270,373	8 9	80.7	11.0
Duluth, Minn.-Superior, Wis	−1,630	371	−.6	30 4	6
Grand Rapids, Mich.	37,630	4,324	15 5	149.0	2 5
Johnstown, Pa	−7,766	704	−2 6	27 7	1.1
Lowell, Mass.	2,835	94	2.2	40.9	.2
Milwaukee, Wis.	90,539	13,623	12 0	141.6	2.7
Racine, Wis	14,171	1,367	15 2	266.5	1 7
Reading, Pa	12,776	1,080	5 3	46.5	1.3
Rochester, N Y.	44,933	4,469	10 3	118 3	1.7
Saginaw, Mich.	17,477	5,570	13 8	154 2	6.0
Sioux City, Iowa	98	192	1	19 0	1.2
Springfield-Holyoke, Mass.	38,749	3,826	10.7	105.3	1.8
Utica-Rome, N Y	19,575	1,524	7.5	142.3	9
Wilkes-Barre-Hazleton, Pa.	−49,286	9	−11 2	.9	.2
Total	771,144	335,537	8 0	84.5	6.6

nonwhite increases noted among the SMA's, about half of the 168 SMA's had increases of from 10 to 49 per cent. In two-thirds of the SMA's the white population experienced a comparable range of increase.

As was pointed out previously, inside the central cities of SMA's the nonwhites increased much more sharply than did the white population over the decade, on the average 48 per cent compared with 10 per cent, respectively. Thus, in 42 central cities the nonwhites doubled compared

TABLE 9

FREQUENCY DISTRIBUTION OF PER CENT CHANGE IN NONWHITE AND IN WHITE POPULATION IN SMA'S, IN CENTRAL CITIES, AND OUTSIDE CENTRAL CITIES, 1940–1950

Per cent increase or decrease	SMA total		In central cities*		Outside central cities	
	White	Nonwhite	White	Nonwhite	White	Nonwhite
Total	168	168	193	193	168	168
−40.0 & over	—	—	—	—	2	4
−20 0 to −39 9	—	3	—	2	1	10
.0 to −19 9	5	6	28	8	10	20
.0 to 9.9	22	18	63	17	12	22
10.0 to 19.9	51	28	35	29	24	16
20.0 to 29.9	29	16	21	14	26	13
30.0 to 39.9	21	20	11	16	21	12
40 0 to 49 9	13	16	8	21	18	10
50.0 to 59.9	12	7	14	11	12	7
60.0 to 69.9	5	7	6	8	12	11
70.0 to 99.9	4	8	—	12	8	5
80 0 to 89 9	1	5	—	5	6	3
90 0 to 99 9	4	3	1	8	4	7
100 0 to 124 9	1	15	4	18	4	6
125.0 to 149 9	—	9	—	9	3	5
150.0 to 174.9	—	3	1	5	3	4
175.0 to 199.9	—	1	—	2	2	—
200.0 and over	—	3	1	8	—	13
Mean	20.0	44.3	10 1	48.3	35.8	32.0
Median	22.1	36.5	11.6	45.0	34.3	29.2

* The 168 SMA's contained 193 central cities, since there were 2 or more central cities in 21 SMA's.

with only 6 for the white. Indeed, in 8 of these central cities, the nonwhites trebled as against only 1 central city for the white At the other end of the distribution, nonwhites in 10 central cities actually declined, whereas the white population declined in 28 central cities. It is apparent also from these figures that the nonwhite and white population changes inside central cities deviated from the average even more widely than did those for the entire SMA's.

It was only in the suburbs that the nonwhite increases from 1940 to 1950 fell somewhat below those of the white population, 32 compared with 36 per cent on the average Here again, however, the population changes among nonwhites varied much more widely than did those for the whites Accordingly, nonwhites actually declined decennially in the suburbs of 34 SMA's compared with only 13 for the white population Moreover, the nonwhites doubled in the suburbs of 28 SMA's compared with only 12 for the white Of these, the nonwhites trebled in the suburbs of 13 SMA's, while the white failed to treble in the suburbs of a single SMA. The SMA suburbs, therefore, experienced a much greater diversity in rate of nonwhite growth than in rate of white growth over the decade

From the foregoing data it is abundantly evident that the rate of both nonwhite and white population growth varies widely among the SMA's, among their central cities, and among their suburbs Such averages as arithmetic means and medians, therefore, fail to characterize accurately the rate of population growth experienced by most SMA's Even the modal rate of change (the percentage by which the largest number of SMA's grew) failed to include either the median or the mean for any of the three nonwhite and three white classifications. The chances are, in fact, that the rate of growth of any specific SMA may fail by a substantial margin to approximate the composite average of the 168 SMA's The frequency distributions shown in Table 9 indicate the wide latitude which is found in the growth patterns of the SMA's.

PROPORTION OF NONWHITES

The variation in percentage of nonwhite population in the 168 SMA's also was very wide, ranging all the way from 0 2 per cent in four SMA's to 45 0 per cent in the Jackson, Mississippi SMA. Although nonwhites comprised almost 10 per cent of all persons living in SMA's in 1950, they amounted to less than 5 per cent of the total in approximately half of the SMA's, as is shown in the frequency distribution of Table 10. This means, of course, that while many of the smaller SMA's

TABLE 10

FREQUENCY DISTRIBUTION OF NONWHITE POPULATION AS A PER CENT OF THE ALL RACE TOTAL IN SMA'S, IN CENTRAL CITIES, AND OUTSIDE CENTRAL CITIES, 1940 AND 1950

Nonwhite as a % of total	SMA total		In central cities*		Outside central cities	
	1950	1940	1950	1940	1950	1940
Total	168	168	193	193	168	168
0 to 4 9	83	91	77	98	111	113
5 0 to 9 9	30	26	41	33	26	15
10 0 to 14 9	19	11	20	14	9	14
15 0 to 19 9	9	9	14	7	6	6
20 0 to 24.9	6	8	9	6	—	3
25 0 to 29 9	7	5	10	10	4	1
30.0 to 34 9	5	4	6	9	6	2
35 0 to 39 9	6	5	10	7	2	3
40 0 to 44 9	2	5	6	7	—	4
45.0 to 49 9	1	2	—	2	1	2
50 0 to 54 9	—	2	—	—	2	2
55 0 to 59.9	—	—	—	—	1	1
60 0 or more	—	—	—	—	—	2
Mean	9 8	8 3	13 0	10 0	5 2	5.4
Median	5 2	4.6	7 4	4 9	3.8	3 7

* The 168 SMA's contained 193 central cities, since there were 2 or more central cities in 21 SMA's.

had a very small proportion of nonwhites, a number of the more populous SMA's had larger than average proportions of nonwhites In no SMA did the nonwhites amount to as much as half of the total population in 1950, although in 21 at least 1 out of every 4 persons was nonwhite, and in 3 SMA's nonwhites comprised over 40 per cent of the all race total

In 1940, nonwhites in SMA's were relatively less numerous, 8.3 per cent of all races, than they were in 1950 In 1940, moreover, well over half of the SMA's had a nonwhite count of less than 5 per cent of the total Yet 9 SMA's in 1940 had a nonwhite population of over 40 per cent of the total. Moreover, in 2 of these SMA's there were relatively more nonwhites in 1940 than in 1950—Jackson, Mississippi and Montgomery, Alabama. Actually, the absolute number of nonwhites in-

creased in both of these SMA's from 1940 to 1950. However, their white population increased so much faster that the proportion of nonwhites declined.

Nonwhites in the 193 central cities of SMA's were relatively more numerous than they were in the SMA totals. Moreover, they increased more sharply from 10 per cent in 1940 to 13 per cent in 1950 on the average. Yet there were only 6 central cities in 1950 in which over 40 per cent of the total population was nonwhite, compared with 9 in 1940. At the other extreme also there was a decline from 98 in 1940 to 77 central cities in 1950 in which nonwhites amounted to less than 5 per cent of all races.

In the suburban parts of the 168 SMA's, the proportion of nonwhites decreased slightly from 1940 to 1950 Of the total, there were 111 in which the nonwhite population amounted to less than 5 per cent of the all race total in 1950—a slight decline from 1940 The number in which nonwhites amounted to from 5 to 10 per cent increased noticeably, however, from 15 in 1940 to 26 in 1950. At the same time, SMA suburban areas with over 40 per cent nonwhite population dropped sharply from 11 in 1940 to only 4 in 1950. As was pointed out previously, however, the decennial changes as between inside central cities and their suburbs were spurious to a minor degree in that there was an expansion of the boundaries of many central cities and a compensating shrinkage of their suburbs.

SMA's WITH SMALL PROPORTION OF NONWHITES

Much recent emphasis has been placed on those SMA's having a large proportion of nonwhite population In many SMA's, however, nonwhites are proportionately very few in number. For example, there were 40 SMA's scattered throughout 18 states in which the nonwhite population comprised less than two per cent of all races in 1950. In 20 of these SMA's, the nonwhites aggregated less than one per cent, as is shown in Table 11. The average nonwhite population for the entire 20 SMA's was only 0 5 per cent of their all race population count.

Although relatively few nonwhites lived in these 20 SMA's, those nonwhites increased by 29 per cent compared with only a 4 per cent increase in the white population from 1940 to 1950. Small though their 0.5 per cent nonwhite proportion was, in 3 of the 20 SMA's the nonwhites doubled over the last decade, and in Manchester the nonwhites trebled. Yet 4 of these SMA's actually lost nonwhite population in amounts ranging from 2 to 37 per cent. In 2 of the 4, Altoona and Scranton, the white population also declined, and in Fall River the white

TABLE 11

THE 20 SMA'S IN WHICH THE NONWHITE POPULATION COMPRISED LESS THAN ONE PER CENT OF THE ALL RACE TOTAL, BY STATE, 1950

Standard metropolitan area	All races 1950	Non-white population 1950	Per cent increase, 1940–50		Non-white as a % of all races
			White	Non-white	
Iowa, Cedar Rapids	104,274	826	16.9	20.9	0.8
Maine, Portland	119,942	465	12 5	17.7	.4
Mass., Brockton	129,428	1,105	8 4	19.8	.9
Fall River	137,298	353	1.8	−37.4	3
Lawrence	125,935	380	.8	30 1	.3
Lowell	133,928	324	2.2	40.9	.2
Worcester	276,336	1,936	9.3	17.3	.7
Mich., Bay City	88,461	380	17 9	53 8	.4
Minn., Duluth-Superior	252,777	1,591	− 6	30 4	.6
N. H., Manchester	88,370	156	7 7	205.9	.2
N. Y., Binghamton	184,698	899	11.4	11 4	5
Utica-Rome	284,262	2,595	7 5	142 3	.9
Pa., Allentown-Bethlehem-Easton	437,824	2,575	10 3	24 9	6
Altoona	139,514	1,150	− 6	− 2 3	.8
Scranton	257,396	818	−14 6	− 5 2	3
Wilkes-Barre-Hazleton	392,241	973	−11 2	.9	.2
S. Dak., Sioux Falls	70,910	426	22 7	82 1	6
Tex., Laredo	56,141	114	22 5	−32 9	2
Wis , Kenosha	75,238	284	18.4	36 5	.4
Madison	169,357	1,059	29.3	124.8	.6
Total	3,524,330	18,409	4.3	28.9	.5

population increased by only 1 8 per cent. With the apparent lack of economic opportunity, in-migration of nonwhites would not be expected.

These SMA's tend to be the smaller ones, with 14 of the 20 having an all race population of less than 200,000. None had a total population of a half million. In absolute terms, 13 of the 20 had fewer than 1,000 nonwhites present in 1950. Of course, none of these 20 SMA's was located in the South.

CAUSES OF VARYING RATES OF SMA GROWTH

A number of factors combine to produce the differences in rate of SMA population growth[15] Since the present analysis has not isolated the separate effect of each of these factors by standardizing for size of SMA, for age of SMA, for per cent of nonwhites, for regional location, for type of industry dominating the SMA, for rate of industrial expansion, etc, only general conclusions can be drawn. Unquestionably, however, the influence of regional location is one of the foremost factors, as is shown in Table 4. Accordingly, western SMA's may be expected to experience a high rate of growth, primarily because of their regional location. Other important causative growth factors are believed to be the age of the central city core of the SMA, with older SMA's growing slower; the economic and industrial expansion and virility of the SMA and of the general area in which it is situated; the climatic, resort, and health appeal of the locality; and to some extent the size of the SMA, although there appear to be differences as between nonwhite and white components, with the nonwhite population growing faster than average in the largest SMA's and with the white growing at about the average rate in the largest SMA's.

Moreover, variations in the tempo of industrial activity account for much of the redistribution of population from 1940 to 1950. Thus, war and postwar expansion in such fields as aircraft, ship building, automotive, light metals, munitions, chemicals, and plastics as well as contractions in such fields as coal mining, railroads, and textiles of some localities, played a real part in the variations in rate of growth of SMA's. After the war, many GI's further accentuated the shifts by failing to return to their former homes and instead by moving to the localities from which their wives came, or to those which they had seen or in which they had been stationed during the war, or to those which promised favorable employment opportunities. The surprising and generally unexpected high continuing levels of postwar employment made this high degree of mobility much easier to achieve than would be expected normally In fact, a recession probably would tend to slow down, but not halt, the population mobility in general and the shift to SMA's in particular

Certain special factors are noteworthy as playing influential roles in the steady influx of nonwhites to SMA's Foremost among these is

[15] For a detailed treatment of growth of all races in SMA's, see Donald J. Bogue, *Population Growth in Standard Metropolitan Areas*, 1900–1950, U S Housing and Home Finance Agency, Washington 25, D. C.

the desire of nonwhites to avail themselves of greater economic opportunity open to them in some localities and regions than in others. This subsumes the opportunity to use and develop their improved training and skills; the freedom to join unions with their promise of equality and a measure of security, without respect to race or color; the legislative protection of fair employment laws and enforcement machinery covering certain cities and States; and the increasing tendency of industries in some areas to employ members of minority groups at all occupational levels on the basis of their competitive abilities Important also are such sociological factors as the search for a larger measure of freedom from segregation and from related social and economic barriers predicated on race or color; and the appeal of urban life and its attendant opportunities to that part of the rural population with a pioneering urge to improve its economic and social life. As these and doubtless other motives impel nonwhites to endeavor to improve their situation, certain traditional routes of migration often are followed Thus, New York is considered a mecca for nonwhites in the South and for Puerto Ricans as well. Similarly, Detroit, Chicago, and St. Louis have long been among the major focal points for nonwhite migration even in the depression years of the Thirties, as have West Coast SMA's during and since World War II. Most Texas SMA's also had substantial inmigration of nonwhites during the Thirties. Many of these new inmigrants remain in their adopted cities, but many others later move on to more distant cities

It is impossible to quantify the effects of each of the many factors which prompt nonwhites to migrate or to remain at home Ordinarily, when a person or a family undertakes as far-reaching an act as moving to a new city, more than one reason underlies that decision. It is clear, however, that the movement of nonwhites to SMA's during the past decade is one of the most dramatic population trends to emerge from the last census, and is continuing today.

SUMMARY

Some of the more salient findings regarding nonwhite population increases in SMA's from 1940 to 1950 follow:

1. As a preface to a consideration of SMA changes, it is emphasized that the nonwhite population in the entire Nation has begun to increase at a faster rate than the white population, in contrast to the century-old precedent of a smaller relative increase in nonwhites.

2. The nonwhite population is being redistributed generally as is evidenced by the fact that nonwhites living in rural farm areas declined

by 30 per cent, while those living in urban areas increased by 50 per cent over the decade.

3. Nonwhites living in SMA's increased over twice as fast as did the white population from 1940 to 1950, or by 44 per cent and 20 per cent, respectively. This is relatively $2\frac{1}{2}$ times as fast as nonwhites increased in all types of areas in the United States.

4. In-migration accounted for an estimated almost two-thirds of the nonwhite population increase in SMA's and in their central cities, and for almost half of their increase in the SMA suburbs.

5. In their movement to SMA's nonwhites have gravitated very sharply to the central cities, whereas the white shift has been markedly to the suburbs. Indeed, in a number of central cities the nonwhites increased, while the white population actually declined. This probably is ascribable partly to the preference of many nonwhites for close-in locations and partly to the lack of new construction available to nonwhites in outlying areas.

6. Regional nonwhite increases were numerically largest in the North Central SMA's and relatively largest in the SMA's of the West. Yet the nonwhites have by no means forsaken the southern SMA's, for the SMA's of the South still embrace the largest number of nonwhites

7. There are great concentrations of nonwhites in a few SMA's. In fact, half of all nonwhites in the 168 SMA's and one-fourth of all nonwhites in the Nation live in 10 SMA's. Nonwhites in these 10 SMA's increased much faster from 1940 to 1950 than did the white population, relatively

8 Yet nonwhites did not increase equally in all SMA's over the decade. In 9 SMA's the nonwhites actually decreased in number, in 31 they doubled, and in 3 they trebled. The white population doubled in only one SMA

9. Nonwhites are not present to the same degree in all SMA's. Thus, in 40 SMA's scattered throughout 18 States nonwhites comprised less than 2 per cent of the all race total in 1950. Still in 14 SMA's, nonwhites amounted to over 30 per cent of the all race total.

10. The reasons for the nonwhite population surge to SMA's cannot be quantified. They encompass such considerations, however, as better employment opportunities, the search for greater freedom from segregation, and the appeal of urban life.

THE PROSPECTS FOR POPULATION FORECASTS

JOHN HAJNAL*
University of Manchester

THE Roman emperor, Constantius, made a law forbidding "anyone to consult a soothsayer, a mathematician, or a forecaster. . . . May curiosity to foretell the future be silenced for ever."[1] However, even the death penalty was, it seems, insufficient to eradicate the condemned practice. Recently, new "scientific," and not infrequently mathematical, techniques of foretelling population growth have provided novel methods, and a new group of experts, for satisfying this basic human need.

Elaborate sets of population projections have in the last twenty-five years become a well established feature of the demographic literature of the Western World. Projections by professional students of population have been increasingly relied on by others—economists, politicians, civil servants—who in former times would have made their own guesses about future population as the occasion arose. Widespread attention was first focussed upon the projections when the prospect of diminishing growth or even decline, which they presented, came as a shock to the public. But the vogue of projections has continued in spite of lack of agreement between the predictions and the facts (which became manifest when the sudden spurt of population growth in the 1940's belied the prophets of doom).

It is the purpose of this paper to argue

(1) that population projections in the future as in the past will often be fairly wide of the mark—as often as simple guesses would be;
(2) that, nevertheless, the frequent preparation of projections will continue;
(3) that a projection can be useful as a piece of analysis even if its accuracy is low;
(4) that simple, unpretentious short term projections should be used to meet most practical needs for population forecasts,
(5) that greater flexibility and variety in techniques for projecting births need to be developed.

* A shortened version of this paper was presented to the World Population Conference held in Rome in September 1954.

[1] *Nemo haruspicem consulat, aut mathematicum, nemo hariolum. . . . Sileat perpetuo divinandi curiositas.* (*Codex Theodosianus*, lib *IX*. tit. *XVI*. l. *4*.)

If some of the following seems unduly dogmatic or provocative, I can only appeal in extenuation to the limitations of space.

Prophecy about the future of human societies is an uncertain business; there seems no reason to expect more success in the prediction of numbers than in the forecasting of other basic features of historical development It is true that population forecasting, like economic extrapolation, is a much more "scientific" and respectable business than, say, predicting the date of outbreak of the next war or the name of the next Pope This impression is, no doubt, in part due to the use of numerical techniques of extrapolation which may suggest analogies with astronomers' calculations of the future position of the stars. Another important factor, however, is the protection given to population forecasters by the slow rate at which population changes take place Usually, almost any method of extrapolation from the past will give results that do not look absurd for a few years (Indeed, widely different methods may often give similar results) This fact enhances the reputation of projectors in another way The consumer of forecasts does not realize that, so far as predictive accuracy is concerned, much of the elaborate technique of forecasters is expended in vain; crude methods could have achieved equally good results Population forecasters are not the only ones who have benefited from this public ignorance. The public opinion poll forecasts of presidential election results have been no better than had the predictions been obtained by assuming at each election that the votes of the major parties were divided in the same proportion as at the previous election [12] In econometric forecasting also, the use of elaborate models has been known to produce worse results than the most naive extrapolation [3].

In the past, prophets have often been upset not so much because their arguments were wrong as because they turned out to be irrelevant. It is the failure of human history to repeat itself, the appearance of the new and the unexpected that renders the search for good methods of forecasting hopeless However much we improve our tools to take care of all that happened in the past, something will sooner or later crop up for which we are unprepared Consider, for example, the forebodings expressed by Malthus and his followers about the British food supply and, consequently, about future population growth in Britain. It was not that they underestimated the possibilities of British agriculture; the argument about its potentialities became irrelevant because importation of food from overseas developed on such a scale that British agriculture declined. Something unsuspected, the development of railways which made practicable the transporta-

tion of food from the interior of the new world, upset their prophecies.

The situation which faced forecasters in the 1930's similarly illustrates how something essentially novel may play havoc with the best judgment based on the past It seems almost impossible that anyone, however great his ingenuity and however extensive his knowledge of the facts, could have foreseen the "baby boom" of the 1940's. Someone who realized what is now known about the possibilities of "bunching" of births (as a result of changes in the pattern in which successive cohorts distribute their births over time) might have taken a more cautious attitude to the forecasts of population decline, but he could hardly have foreseen birth rates as high as those of the 1940's Indeed, suppose a man had had the gift of prescience about everything in the 1940's except the number of births. He would have known about the war and about the economic prosperity. Even if he had in addition been the most competent demographer in the world, he would almost certainly have guessed wrong He would probably have argued, for example, that in wartime Britain, groaning under austerity and exposed to German bombardment, the number of births would (in spite of full employment) be very low. If with all our hindsight we cannot blame the demographers of the 1930's, what reason have we to expect better luck in the future?

The factors whose effects on future growth we can calculate are likely to be frequently outweighed by the unpredictable. It is this which accounts for the failure of more complex techniques to yield more accurate results than simple techniques and which casts doubt on the value of forecasting We cannot hope to develop better methods which yield forecasts clustering more and more closely round the true future population New and more complex techniques which may yet be invented are, I think, just as liable as past techniques to be fairly often upset by the unpredictability of history. They will probably just as often—and that means rather frequently—give results which are very wide of the mark and less accurate than crude guessing.

This view of the situation is confirmed in the field of local forecasting, i e , forecasting the population of smaller geographical units within nations National forecasts are not often repeated by identical methods in identical circumstances However, in local forecasting much more experience has accumulated, particularly in the United States, by which the success of different methods may be judged by a fair number of trials in comparable circumstances (for example, when forecasts for each of the forty-eight states are made) Series of forecasts have been specifically computed for such comparisons The fol-

lowing quotation from a survey of the American experience by Siegel illustrates the point I wish to make: "From the point of view of accuracy, the simple methods appear to be just as accurate as the complex ones. . . . Average errors are not always disturbingly large for short-term forecasts, but extremely large errors may appear in particular cases from the very start with any of the methods tested. The tests are consistent in showing an increase in the error with an increase in the length of the forecast period; 20-year forecasts generally have large errors and a sizeable proportion of errors exceeding 10 per cent" [16].

One argument from the experience of population forecasting may be urged against the view here taken that future events are essentially unpredictable Projecting the survivors of cohorts already born—e g., for projections of the labor force—has generally proved reasonably successful. However, it may be doubted whether success in this field really amounts to successful forecasting of future events. Projected mortality rates and forecasts of deaths have in fact often been very inaccurate. However, with the low death-rates now current in Western countries, a large error in projecting deaths corresponds to only a small proportionate error in the survivors (except at advanced ages). To predict the population aged 20 five years hence amounts essentially to noting that those now aged 15 will then be 20 This sort of calculation (which may involve technical difficulties, for example owing to errors of enumeration) is very useful indeed, but it is hardly the prediction of anything which is seriously in doubt Of course, the procedure involves the prediction, or assumption, that there will be no catastrophic mortality (say, owing to H-bombs). But, to the making of this assumption the special techniques of demography are not relevant.

Are population projections still produced and used by the public only because knowledge of their failure in the past has not been widely diffused outside the circle of demographers?[2] Or because experts, who are aware of the fate of past forecasts, are nevertheless still hopeful that better methods may be found? It seems doubtful whether a radical pessimism about the possibility of forecasting has much chance of general acceptance in the present atmosphere of confident expansion in the social sciences But even if such a view were to spread it would not mean the abandonment of all population projections.

The demand for guesses at future population seems to have been

[2] So far as I know, efforts to publicize the failure of population projections have been made only in the United States Professor Joseph S. Davis has published some scathing comments with this aim. (See, for example, reference [4]).

growing. The increase in government planning and the growth of social insurance and social services is perhaps the main reason At the same time, the statistical material available for the study of population and the techniques used for their analysis have become much more plentiful and complex. Increasingly they require specialists to handle them. It is improbable that detailed analysis of the factors at work in population growth and guessing on the basis of that analysis will be given up in favor of the older practice of guessing, without analysis, on the basis of total population figures and simple rates of growth. Analysis will not protect us against unforeseen developments, but it may often give useful warning against uncritical extrapolation of past experience. Moreover, the need in modern times is often not for forecasts of total population but for projections of special classes, such as particular age groups, numbers of people with certain numbers of children, etc., and such projections generally necessitate the use of specialist techniques.

It seems unlikely that the number of whole or part-time population specialists will drop or that they will cease to turn out population forecasts for a variety of purposes The situation in economics offers something of a parallel The economists demonstrated conspicuous failure in prediction much earlier than the demographers—at the time of the great depression Yet the number of economists employed in industry, government and international bodies has risen prodigiously since the slump. The rapid expansion in government activities and in the volume of economic statistics has no doubt helped to bring this about An analogous situation exists, on a smaller scale, in demography Population forecasters have no need to fear that failure will result in loss of jobs.

The demand for population forecasts is in part nourished by motives which are slow to react to evidence about their inaccuracy. Even very inaccurate forecasts often meet a need, the same age-old need which (perhaps even more than curiosity) has caused people to turn to forecasters or soothsayers of all kinds, the need to take decisions. This seems to be the meaning of the demand so often faced by the statistician "Give me a forecast, any figure is better than none."

It is sensible to bring to bear on a decision whatever scientific apparatus is available. There is, however, some danger in this process. Anyone in the early 1940's having to plan the educational facilities for children of elementary school age in ten years' time would, as we now know, have acted more wisely if, instead of looking at population projections, he had assumed that nothing was known about the future

numbers of births, that an increase was as likely as a decrease, and that plans must be prepared for both eventualities However, decisions in government and business must usually be based on some view of likely future development and we can assume that population projections will continue to be used for this purpose.

What then can be done towards better population forecasting? We might begin by assessing what went wrong with past forecasts A striking indictment could be drawn by compiling examples of wide discrepancies[3] between populations as forecast and as enumerated. A second line of criticism, which was mentioned earlier, might be that the complex exertions of the forecasters achieved predictions which were often further removed from the facts than naive extrapolations (based on, say, the growth of the 5 years before the forecast was made). These criticisms concern the *accuracy* of the projections However, there is yet another question The authors of projections have in any case usually guarded themselves against attack on grounds of accuracy by statements to the effect that projections are not predictions, but only show the results of certain hypothetical assumptions The question is: were the assumptions relevant? Though their authors cannot be blamed for it, the projections of the 1930's and early 1940's turned out to be wrong in *analysis*. They were intended principally to demonstrate one thing, the prospect of the end of population growth and of actual decline in the near future. If the populations of Western nations today fell short of their predicted numbers by as great a percentage as they in fact exceed them, this would probably be considered a triumphant justification of the analysis underlying the projections. It could also be considered as confirmation of the analysis if, though the populations are now larger than predicted, decline had been averted only by drastic measures to raise the birth rate (this is widely believed to have been the case in France). The success of a population projection as a piece of analysis is not measured by the percentage difference between the projected and the actual populations.

Perhaps the greatest achievement so far in the field of population forecasting was a projection constructed by Edwin Cannan [2] in 1895. At a time when the population of England and Wales was still growing by more than 10 per cent per decade with no sign of slacken-

[3] The relation to the facts of projections of the U. S population has been extensively examined by Davis [4] and Dorn [5] That the situation in Western Europe is equally unsatisfactory may easily be seen, for example, by studying the collections of projections for various countries by different authors which were assembled by Glass [7] and Notestein and others [13] The latter volume contains an Appendix which presents by means of graphs the results of earlier forecasts for the individual countries of Europe. It also contains a full bibliography of earlier projections.

ing, and population forecasts were commonly based on the assumption that this rate would continue, Cannan predicted that growth would cease in the 20th century. He used essentially modern methods. He diminished the population of each age group by allowance for mortality and added an allowance for births based on the ratio of births to persons of reproductive age He arrived at a prediction which was at variance with commonly accepted ideas of his time[4] and foresaw an important development in population growth by an analysis of the factors at work. As a result he could publish 36 years later, in 1931, a paper which said in effect "I told you so" [1]

Yet the accuracy of his forecast was not outstanding By 1911, only fifteen years after the forecast was made, the population enumerated at the census exceeded the prediction by 7 per cent and by 1916 the population had increased beyond his estimate of the maximum it would ever reach[5]

Cannan's work could not be regarded as useful from the point of view of the practical consumer of forecasts in his day. He condemned, as having no rational foundation, the extrapolation of the growth rate of the last intercensal period. However, if one carries the population forward for two decades from the 1891 census, using the 1881–91 growth rate of 11.65 per cent, one obtains for 1911 a figure which differs from the census count by less than one quarter of 1 per cent.

Projections of greater accuracy than Cannan's are not rare in the history of population forecasting. This can easily be illustrated from the survey of population projections for the United States which was made by Dorn [5] in 1950. He gives many comparisons between actual and projected populations Particularly instructive instances are provided by two projections published in the 1920's The first, by Pearl and Reed, was constructed by fitting a logistic curve to the census counts of 1790 to 1910. The projections came within one per cent of of the 1920 and 1930 census totals, overestimated the 1940 population by 3.5 per cent and underestimated the 1950 population by about 1 per cent.

The second projection was published by Whelpton in 1928 and was

[4] He wrote for example· "During the last twenty years most of us have not succeeded in detecting any considerable change in the manners and customs and practices which affect natality, and yet it only requires a continuance of the change which has undoubtedly been going on to bring about a state of things which could cause the possibility of a decline in population, instead of the possibility of over-population, to be the bugbear of alarmists " (This was written in 1895!)

[5] He did not in his 1895 paper actually give any projected figures except for the maximum population I have deduced his forecast for 1911 from the graph he gives He had no illusions about the value of working out precise figures of future populations, "The value of the diagram lies not in its prediction of a maximum population of thirty-seven millions, but in the fact that it shows how a cessation of growth may be reached within no very long period without any violent or unnatural changes."

made by the modern component method. It gave results very similar to the Pearl-Reed logistic. The total population predicted exceeded the census figure by five per cent in 1940, but for 1950 was within 1 per cent of the census count.[6] But these authors themselves proved that their success was only accidental. As a consequence of the rapid fall in the birth rate they came to feel that their projections were too high. Whelpton, and later Pearl and Reed, issued revised estimates which were lower than the original ones and turned out to be in worse agreement with the facts.[7]

The sort of achievement which Cannan accomplished in 1895 is, of necessity, rare. It can only occur by a meeting of the opportunity and the prepared mind. Its function is not to produce a confident prediction (this I believe to be impossible), but to instill skepticism about other people's confident predictions, to reveal by detailed analysis that the implications of the present statistics have been misinterpreted, to show that by "continuance" of what has been going on and "without any violent or unnatural changes" (in Cannan's words) the future may turn out to be very different from what is commonly supposed. This is achievement of a high order.

At the same time, this is not the kind of excellence which is required by most consumers of population projections. What they need is forecasts (generally relatively short-term forecasts) which differ by only a small percentage from the actual population, whether they be based on a correct appreciation of the forces at work or on black magic. The elaborate sets of projections often prepared by demographers are not well suited to this need, not only because they are not very accurate, but because, owing to the amount of work which their preparation involves, they are often a year or two behind the latest data by the time they are published and cannot very well be kept up-to-date. This can be a serious disadvantage in a period of sharp fluctuations.

In recent years a person needing a population forecast for, say, five years ahead, would often have done better to use the latest data and make a crude guess, rather than rely on a population projection published a few months earlier. Of course, the non-specialist will often find it inconvenient to find his way about population statistics, and one of the services which demographers or a central statistical agency can perform is to supply, at regular intervals, guesses about future popula-

[6] The agreement was much less close for individual age groups.

[7] Pearl and Reed, however, in publishing their revised figures, "did not express a clear choice" between their two logistics though they "seemed to have a slight preference" for the later one. The logistic fitted to the counts for 1790 to 1910 turned out to be a far better predictor of the 1950 figure than the curve which also took into account points for 1920, 1930 and 1940!

tion based on up-to-date information. In Western countries annual estimates of population by sex and age are now generally prepared as a matter of routine. It is only necessary to use the latest estimate to calculate "survivors" for a few years ahead, and add in a guess for births (and migration if desired). Both the survival rates and the birth estimates can be of the crudest kind, but should not be at variance with the latest information. Thus the survival rates should not be based on an official life table five years old. (In recent times the decline in mortality has often exceeded projected mortality declines very soon after the projection was made.) There is no reason why the figure for births should not be the total for the last five years or something equally simple. Quick, crude methods can, of course, be employed also to produce a range of predictions rather than a single "best guess."

In some circumstances, the fitting of simple growth curves[8] will be useful. On the whole, however, growth curves are most likely to find application in under-developed countries with poor statistics. Projection under these conditions is a separate subject outside the scope of this paper.

In spite of the fate of most of the projections done in the period 1930–1945, sets of quite elaborate projections have continued to be produced in recent years. In most cases the technique of computation is the same as that used mainly in the 1930's, i.e., to start with a base population divided by age and sex, work out the number of survivors by means of survival ratios based on age-specific mortality rates and add in births by applying age-specific fertility rates to the female population. (Generally the calculation is carried out with rates specific by 5-year age groups, yielding future population figures at 5-year intervals.)

We may begin by asking why this technique is an improvement over the older methods of fitting curves to total population, or extrapolating rates of growth or crude birth and death rates. The standard modern technique has, I think, two main advantages, which are related to each other. The first is that this method reveals certain future developments which are "inherent" in the present age distribution. It is this

[8] However, the special prominence of the logistic curve in this field seems exaggerated. (a) Some writers appear to believe that an S-shaped curve is necessarily a logistic. In fact, even the biological data on which the reputation of the logistic is partly based, can be equally closely fitted by other S-shaped curves with the same number of arbitrary parameters [6]. (b) Fitting a logistic to a set of data is not a determinate process. It is often possible to fit two different logistics to a given set of figures such that both give a very good fit and yet they give very different results when extrapolated. This is true even when several observations on both sides of the point of inflection are available (see, for example, reference [15] particularly p. 35). When, as is so often the case in demography, all the observed figures lie on one side of the point of inflection, a wide variety of predictions can generally be produced by fitting a logistic.

point which largely explains the great significance which has been attached to this method of projection Analysis of the effects of the age distribution on the crude birth and death rates made possible the predictions of population decline which captured the public imagination. The second advantage consists in the correspondence which is believed to exist between age-specific mortality and fertility rates and the causal factors about which forecasters feel they can prognosticate For example, suppose that a person forecasting the population of England 30 years ago had reasoned correctly that mortality would probably fall greatly, owing to increases in the general standard of life, the development of medical services, etc. If he had concluded that the crude death rate would fall he would have been wrong. While age specific death rates have fallen dramatically the crude death rate has hardly changed owing to shifts in the age structure.

These are considerable advantages. Yet the usual manner of forecasting births in more elaborate projections hardly seems adequate to present day needs. No doubt the number of births is influenced by the number of women of child-bearing age; for example, an upper limit is thus set to the number of births that can occur. But, within the range of variation which it is of interest to predict, the relationship between changes in these two factors over time is not close. For example, in England (and some other countries) the recent "baby boom," which occurred in a time when the number of women aged 20–40 was beginning to fall, followed upon a period when an increasing number of women was accompanied by a declining number of births It is very strange that so many forecasters have expended great computational labor on taking account of the effect on future numbers of births of the number of women in each five year age group, while entirely neglecting other factors on which statistical information was readily at hand Such factors are: impending change in the ratio of men to women, abnormal weighting of the married population with recent marriages, sharp diminution in single population leading one to expect a decline in the number of marriages, large fluctuations in the distribution of births by parity in the last few years, etc. Even crude calculations to illustrate the effect which those various influences might have on future numbers of births seem more worthwhile than elaborate extrapolations of age specific fertility rates.

The justification for preparing complex population projections must be the expectation of following in Cannan's footsteps The forecaster must be inspired by hope of analytical insight rather than of accurate

prediction and practical help to administrators. The technique used must be appropriate to this aim.

Techniques of projecting into the future are, whether by design or not, intimately connected with methods of analyzing the past The standard technique of projecting births arose out of the habit of analyzing fertility in terms of the age-specific rates of women, which were added up into gross and net reproduction rates. It is not possible today to frame sensible assumptions about future fertility in terms of the traditional age-specific rates or gross reproduction rates. Straightforward extrapolation of the past trend—a long-term decline followed by sharp fluctuations—is meaningless. One may try to draw inferences about the factors influencing people to want more or fewer children, but the experience of the past has shown that it would be just as wrong to pass from the belief that "people will want fewer children" to the conclusion that "the gross reproduction rate will fall," as it would be to predict a fall in the crude death rate from an improvement in health conditions. To put the same point in another way, to assume maintenance of current age-specific fertility rates may be to imply very strange assumptions about the size of family, in the same way as continuance of the crude death rate may involve unlikely and unintended assumptions about mortality.

Some pre-war projections went beyond the simple age-specific fertility rates in forecasting the number of births—for example, by making adjustments for variations in the ratio of men to women or projecting the proportions of women married and applying separate fertility rates to married and unmarried women.[9] Recently the efforts at applying novel techniques for analyzing the number of births have occasionally found expression in forecasting Thus births have been projected by forecasting the number of marriages and then forecasting the births from the marriages [14, 10, and 9] The procedure of basing the forecast of marriages on the population of single persons has commended itself, since in recent years there have been sharp decreases in the proportions remaining single in various age groups, which suggests that a fall in the number of marriages is in prospect. To pass from marriages to births, assumptions may be made concerning (a) the average number of births to be produced by each marriage and (b) the distribution of these births over the years subsequent to the marriage.

[9] This is, however, often an undesirable method of taking account of variations in marriage patterns in projections, as also in computing reproduction rates (the reasons are discussed, in the latter context, in [8] para. 128).

The procedure may be varied by projecting first births from marriages, second births from first births and so forth, thus using directly the known numbers of first, second, etc. births, as well as of marriages, in the recent past. This requires assumptions concerning (a) the number of births of order $n+1$ per birth of order n and (b) the distribution of the births of order $n+1$ over the years subsequent to the occurrence of the births of order n. Another procedure, following a different line of approach, is to use the traditional age-specific female fertility rates, but arrange them in terms of successive cohorts and make assumptions about the total number of births of cohorts Here again births of different order may be treated separately (see [17] which is summarized in [11]).

Many other procedures, ranging from simple methods such as projecting births on the basis of the number of men, to complex computations, are implicit in modern methods of analyzing fertility In one sense, work on these lines will make the interpretation of forecasts more bewildering. New techniques may suggest the possibility of novel patterns of population change different from those predicted by traditional methods For example, where two projections have been made by the traditional method, using different fertility assumptions, the difference between the resulting numbers of births is slight at first and then increases indefinitely, the direction of the difference being constant If projections are made using the cohort principle[10] and varying the distribution of births over the lifetime of different cohorts, this will no longer hold and quite natural assumptions may result in completely different patterns of variation Projected births may differ widely even for the first few years Moreover, it may be that on one assumption the births will be lower in the first time period, but higher in the second time period, than under another assumption as to fertility behavior. Such patterns may make the estimation of maximum and minimum figures within which the population will lie a complex problem from the purely technical point of view.

The applicability of techniques of forecasting depends, of course, in part on their complexity and on the availability of the statistical data. Some methods require materials which are only rarely available. Principally, however, the interest of a method depends upon what has been going on in the population to which it is to be applied. Analysis may reveal features in the recent history and structure of the popula-

[10] By cohort principle, I mean a method which may make more births in one time period result in fewer births in the next, i.e., which allows for shifts in the distribution of births over the lifetime of successive cohorts. In this sense all the methods discussed in the previous paragraph embody something of the cohort principle.

tion which can be expected to have effects on future births, effects which are not adequately taken into account by traditional methods. The problem of the availability of data thus tends to solve itself. If there are data to show the effect of a certain change on births in the recent past, these data can almost always be used for a calculation concerning the future.

It would be pointless to try to find the best method or the right method of projecting births, just as, I believe, it is pointless to look for the right method of measuring fertility. In fact the two are the same thing; a technique of projection implies a technique of measurement and vice versa. Moreover, the need is not for elaborate projections by this or that particular new method. Rather each forecaster should try to use the information available to him to the best advantage to throw light on possible future development. This may involve trying several methods, if necessary by rather crude computations. In addition, it would be desirable to have some special studies especially to compare the properties of various techniques of projection and to devise short-cut methods of calculation, so that the forecaster may have a better view of the various tools at his disposal and be able to apply them without prohibitive expenditure of time.

If there is a general lesson to be drawn from all this, it is, I think, first that as little forecasting as possible should be done, and second that, if a forecast (more elaborate than the quick calculation discussed earlier) is undertaken, it should involve less computation and more cogitation than has generally been applied. Forecasts should flow from the analysis of the past. Anyone who has not bothered with analysis should not forecast. The labor spent in doing elaborate projections on a variety of assumptions by a ready-made technique would often be much better employed in a study of the past Out of such study may occasionally come important insights about unexpected possibilities in the future.

REFERENCES

[1] Cannan, Edwin, "The changed outlook in regard to population, 1831–1931," *Economic Journal*, 41 (1931), 519–34.

[2] Cannan, Edwin, "The probability of a cessation of the growth of population in England and Wales during the next century," *Economic Journal*, 5 (1895), 505–16.

[3] Christ, Carl, "A test of an econometric model for the United States, 1921–1947," *Conference on Business Cycles*. New York: National Bureau of Economic Research (1951), 35–106.

[4] Davis, Joseph S., *The Population Upsurge in the United States*, War-peace Pamphlets, No. 12. Food Research Institute, Stanford: Stanford University, 1949.

[5] Dorn, Harold, "Pitfalls in population forecasts and projections," *Journal of the American Statistical Association*, 45 (1950), 311–34

[6] Feller, W., "On the logistic law of growth and its empirical verification in biology," *Acta Biotheoretica*, 5 (1940), 51–65.

[7] Glass, D V., *Population Policies and Movements*. Oxford: Clarendon Press, 1940.

[8] Hajnal, J., "Births, marriages and reproductivity in England and Wales," *Reports and Selected Papers of the Statistical Committee*, London Papers of the Royal Commission on Population, II (1950), 303–422.

[9] Henry, Louis, *Fécondité des Mariages*. Paris: Institut National d'Études Demographiques, Travaux et Documents, Cahier No. 16, 1953

[10] Hopkin, W A B., "Population projections for Great Britain 1947–2047," *Reports and Selected Papers of the Statistics Committee*, London: Papers of the Royal Commission on Population, II (1950), 213–301.

[11] Kiser, Clyde V , "Fertility trends and differentials in the United States," *Journal of the American Statistical Association*, 47 (1952), 25–48.

[12] Mosteller, Frederick, "Measuring the error," *The Pre-election Polls of 1948*. Report to the Committee on Analysis of pre-election Polls and Forecasts, Bulletin 60 New York: Social Science Research Council (1949), 54–80

[13] Notestein, F W., and others, *The Future Population of Europe and the Soviet Union*. Geneva: League of Nations, 1944.

[14] Quensel, C E., "Om befolknigsprognoser med särskild hänsyn till aktenskapsfrekvens, familjebildning och erforderlig familjestorlek," *Statistiska undersokningar kring befolkningsfrågan utforda av 1941 års befolkningsutredning* Sweden Statens Offentliga Utredningar, 53 (1945), 223–39

[15] Rhodes, E C , "Secular changes in death rates," *Journal of the Royal Statistical Society*, 104 (1941), 15–42

[16] Siegel, Jacob S , "Some aspects of the methodology of population forecasts for geographic subdivisions of countries " Paper submitted to the World Population Conference, Rome, September 1954 To be published

[17] Whelpton, P K , *Cohort Fertility Native White Women in the United States* Princeton: Princeton University Press, 1954.

A TECHNIQUE FOR ESTIMATING THE POPULATION OF COUNTIES

HUGH H. BROWN
California Taxpayers' Association

CURRENT estimates of the population of counties are desired for a number of purposes which, however, are so widely recognized, that discussion of them seems superfluous. Several methods have been developed for making postcensal estimates of county populations [6, 7, 9, 13] This article outlines another which may be applied when certain statistical data are available, most important of which are school enrolments by grade and births by place of residence. Also desired are deaths by place of residence, voter registrations, and automobile registrations The estimating procedure derives from the logic of the problem and the materials.

TERMS AND FORMULA

The basic statistical series used for current population estimates is the count of children in grades one through eight. This is identified for formula purposes as the school count. The ratio between the total population of a county and the number of children in grade school at the time of the 1950 Census is easily calculated, and is referred to as the Census ratio. Unfortunately, that ratio is constantly changing as the result of variations six to fourteen years earlier in the birth rate together with modifications resulting from migration, economic activities, and other influences Consequently, to produce current estimates of population based on elementary grade counts, it is necessary to introduce a third element—a correction factor.

The theoretical formula is, then school count$\times$Census ratio$\times$correction factor$=$population estimate

In practice, the correction factor is composed of two factors to which, for identification, have been given the names of "ratio change factor" and "conversion factor." The ratio change factor compensates for changes in the Census ratio. The conversion factor corrects for an error in the ratio change factor, thus "converting" it to a correction factor Lastly, certain adjustments are made as the result of specific tests. In its full statement, then, the formula is: school count $\times$Census ratio $\times$ratio change factor$\times$conversion factor$\pm$adjustments$=$population estimate.

How each of these components is derived and used is explained in the following paragraphs.

SCHOOL COUNT

California public schools count active enrollees as of each March 31 and October 31. Counts are collected and tabulated by the State Department of Education and are published in their monthly bulletin, "California Schools." The published summaries show totals by county by grade by sex from kindergarten through grade 14 and in special classes.

In early 1950, the California Taxpayers' Association canvassed church organizations to find those operating regular parochial day schools and boarding schools Those identified report to the Association their end-of-school year (June) and beginning-of-school year (September) counts which the Association summarizes and releases by county by grade from kindergarten through grade 14

For population estimating purposes, parochial and public school counts for grades one through eight are combined, including elementary level classes for ungraded, physically handicapped, and mentally retarded pupils. The parochial schools follow the State Department of Education's manual of rules and regulations controlling age of entry and other requirements. Essentially, then, the school counts provide a statistical series which is consistent from county to county and year to year. Attendance at school by children in the elementary school ages is virtually at saturation in California though there is a small range of difference between industrial and agricultural areas, attendance being less complete in the latter. From 1940 to 1950, there was some improvement in completeness of attendance and there may be more, particularly in the agricultural areas, between 1950 and 1960, but it will probably be slight. It is assumed, for the present, that the character of each county with respect to completeness of attendance is stable.

One inconsistency has already occurred since 1950, however. The State legislature changed the age minimum for entrance into the first grade from 5 years 6 months to 5 years 9 months, effective in the fall of 1952. It is necessary to estimate the number of students excluded by the new rule. The estimate is made by reference to the corresponding births.

In order to provide the school count in the form needed for current population estimates as of each January 1, the total counts of March and October are projected two months to January 1 of the next calendar year. This leads to a slight overstatement as there is a tendency for children to be held back during the spring and summer months and entered in a wave in the fall. Various other factors, of which seasonal

agricultural migration is the most important, also affect the two counts. Each year a preliminary school count is produced by the March through October projection and each following year a revision is produced by differencing to January between the prior October and the new March counts. This is more accurate and minimizes distortions due to seasonal agricultural migration and the tendency to hold new students back for the fall session. Preliminary population estimates based on the two-months projection are revised the following year according to the revised school count

Several years of experience in relating the projected to the revised school counts now make possible the determination of adjustments for seasonal variations in counties where the pattern is fairly stable. These are applied to each new projection of the preliminary count to make it more closely approximate what the later revision will be.

CENSUS RATIO

The Census ratios were calculated by dividing the 1950 Census population of each county by the 1950 spring school counts. These ratios were studied by arraying them from smallest to largest and differentiating them into six groups They range from Madera County with a ratio of 5 58031 to San Francisco with a ratio of 12.49347. This array gives a basic distribution for a type of testing called "pattern testing." The six groupings are arbitrary but are based on demographic and economic characteristics of the counties Group 1 consists of the four major urban-industrial counties Groups 2 and 3 are semi-urban counties with mixed industrial and agricultural economies, Group 2 being in the San Francisco Bay Area and Group 3 in Southern California. Group 4 is Sacramento Valley counties, and Group 5 is San Joaquin Valley counties, all primarily agricultural but with important industrial developments in spots Group 6 is all others

At the top of the table with the lowest ratios are the counties with *relatively* the greatest number of children in the elementary school ages while at the bottom of the table are those with the least. This basic characteristic derives, of course, from the age distribution at the time of the Census, which derives from the level of the birth rate six to fourteen years earlier, which, in turn, derives from the economic and social character of the county. The top of the table is the extreme of ruralness, agriculturalism, and youthfulness of the population while the bottom is the extreme of urbanness, industrialization, and aging of the population.

The way the groups fall in this demographic array, the sequence of

the counties within each group, and the interlacing of the groups provide patterns which are of importance in spotting changes in demographic character when "pattern testing" is applied to other materials used in the estimating process The research statistician becomes familiar with these patterns and can recognize deviations from them or variations of them. In terms of principle, where there are no *absolute* yardsticks or standards of what this or that demographic characteristic should be, significant information may be derived *on a relative basis* from interrelationships among the counties

RATIO CHANGE FACTOR

The Census ratios—which, again, are the ratio of the total population to pupils in elementary school grades in April, 1950—may be said to measure the relative density of children in that age bracket within the county populations If the ratios remained constant, population estimating based on school counts would be simple, but, as already stated, they do not. They change continuously and "ratio change factors" are necessary to compensate

The ratio change factors are produced from birth rates. The birth rate of a specific year reflects the relative number of children born into the population in that year and that relative relationship continues sufficiently unchanged during the six to fourteen years the children are in elementary school to make its use in this application practicable For working purposes, the children who were in school in April, 1950, are considered as having been 6 to 14 years old and as having been born during the years 1936 to 1944, inclusive This is not exactly true but it is practical since it permits use of annual totals of births and precludes the necessity for totaling births by 12-month periods including parts of calendar years. Moreover, errors resulting from this inaccuracy are compensated for in ways still to be explained The age span of six to fourteen years is used because it gave better results when checked against the 1940 Census data than the six to thirteen year spread. Consequently, it was adopted for use during the 1950–60 decade.

For each school year, the birth rates for the nine years six to fourteen years earlier are added together to give a specific sum for that year. Thus, adding the birth rates together for 1936 to 1944, inclusive, gives a total of 152 50 for July, 1950. The birth rates are considered to be as of July 1 since they are computed from place-of-residence births for each calendar year and an estimate of population as of July 1. The purpose of these figures will become apparent by reference to Table 1. Note the total of 152 50 for July, 1950 in column 3. This is the total of the birth rates for 1936 through 1944 as shown in column 2.

TABLE 1
CUMULATED BIRTH RATES, CALIFORNIA

	Birth Rate		Cumulated Rates, 6–14 Years Earlier			
					Ratio*	
	Year	Rate	Years Included	Cumulated Rates	1940 Base	1950 Base
	(1)	(2)	(3)	(4)	(5)	(6)
	1925	17.60				
	1926	16.34				
	1927	16.32				
	1928	15 62				
	1929	14 77				
	1930	14.83				
	1931	14 07				
	1932	13 28				
	1933	12.67				
	1934	13.04				
	1935	13 18				
	1936	13.51				
	1937	14 67				
	1938	15.34				
	1939	15 26	1925–33	135.50		
(April)	1940			132 08	*1.00000*	1.14173
(July)	1940	16 10	1926–34	130.94	1.00871	1 15167
	1941	17 27	1927–35	127.78	1 03365	1.18015
	1942	19 92	1928–36	124.97	1.05689	1.20669
	1943	20.45	1929–37	124.02	1.06499	1.21593
	1944	19.98	1930–38	124.59	1 06012	1 21037
	1945	19 50	1931–39	125 02	1 05647	1 20621
	1946	22 68	1932–40	127.05	1 03959	1 18693
	1947	24 87	1933–41	131 04	1.00794	1.15079
	1948	23.88	1934–42	138.29	.95509	1.09046
	1949	23.68	1935–43	145.70	.90652	1.03500
(April)	1950			150 80	.87586	*1 00000*
(July)	1950		1936–44	152.50		.98885
	1951		1937–45	158.49		.95148
	1952		1938–46	166.50		90571
	1953		1939–47	176.03		.85667
	1954		1940–48	184 65		.81668
	1955		1941–49	192 23		.78448

* Base period cumulative rates divided by those of each year.

We may regard the sum of 152.50 as an *indicator of relative density of children of elementary grade ages in the population of California as of July, 1950.* The sum for July, 1949 is 145 70. Proportioning the difference between the 1949 and 1950 figures gives a sum of 150.80 as of April 1, 1950, the date of the Census This, then, is the *base* for the 1950–60 decade Note next that the sum for 1954 is 184 65. This demonstrates that the relative number of children of grade school age in the population was substantially greater in 1954 than it was in 1950. For population estimating purposes, it means that the 1950 Census ratio must be reduced to compensate for the change in relative density of grade school children. The reduction is made by calculating the ratio of change to each succeeding year from the base year and multiplying the product of school count times Census ratio by its reciprocal. For working purposes, we skip one division and calculate the ratio of the base year to each succeeding year directly. Thus, 150.80/184.65 =.81668. Consequently, estimates made by multiplying the 1954 school count by the 1950 Census ratio must also be multiplied by .81668 to correct for change in the ratio between population and grade school count.

The series with 1950 as the base year for use during the current decade is shown in column 6. The density of children of grade school ages within the population is now rising, as we know it must because of the much publicized "birth wave." If the birth rates, the cumulated rates, and the ratios to the base period are all plotted on the same chart, these relationships and the shapes of the curves become apparent.

Census ratios and the cumulated birth rates are *both* measures of the relative density of children of elementary school age within the population The Census ratios are one-time items, however, while the cumulated rates are a continuous series It is possible, therefore, using the cumulated rates series, to calculate the ratio between a base year and any specific year and the reciprocal of this ratio, is then, an *index of change of density* which can be applied to the Census ratio.

It is clear that the ratio change factor must be a multiplier which can vary above and below unity, so that in a phase following increasing birth rates, such as the present, it can decline and fall below unity while in a phase following decreasing birth rates, such as during the 1940's, it can rise and become greater than unity. The ratio to base period of the cumulative birth rates meets this requirement. It rises and falls pursuant to birth experience six to fourteen years earlier.

Ratios with 1940 as the base (actually in reciprocal form, of course) are shown in column 5. The cumulated rates as of April, 1940 were

132.08. In 1943, the density of children was at its lowest, the sum was 124.02 and the corresponding change factor was 1.06499. After 1943, this multiplier diminished until it passed through unity (the point of zero change) in 1948, and for April, 1950, it was .87586. Now, if we were to check the accuracy of this method of estimating the population of each county against the 1950 Census by using 1950 school counts, the 1940 Census ratios and correction factors, the county counterparts of this statewide factor would be the correction factors to use. Analysis indicates that these ratios do, in fact, come very close to being what such correction factors should be, but it is impossible to make a conclusive study because the method of counting children in school was changed in 1947. Nevertheless, analysis demonstrates that ratios calculated from cumulated birth rates come close enough for practical purposes.

The series indicated in column 6 of Table 1 has been calculated for all of the 58 counties to produce a ratio change factor for each county each year beginning with 1950 Ratios for 1954 are shown on Table 2 which distributes the counties in groups as mentioned earlier under the term of "pattern testing." The table, in essence, shows relative change in the Census ratios from 1950 to 1954. Every county has its own characteristics, its own relative density of children and, hence, its own degree of change. As with the Census ratios, there is a wide range from highest ratio change factor to lowest.

It is of interest to note that the more urban type of county runs at the bottom of the scale on Table 2 while the more rural type runs toward the top. This means that the Census ratios of the urban-industrial type counties must be more sharply discounted than those in the rural-agricultural type because the increase in density of children due to the rise in the birth rate from its previous levels is relatively greater in the urban counties than in the rural ones

In using the ratio change factors, it is necessary to make the underlying assumption that the people who come into a county through migration are demographically similar to those already there. It is thus assumed that city people migrate to San Francisco and farm people to Madera county. While this will usually be true, it is not true always and the demographic character of a county may change This fact is recognized and necessary adjustments are provided in testing procedures which will be explained presently.

In dealing with materials of this type, one must be constantly aware that any specific item for any county in any year may be off-normal. The ratio change factor for a particular county may tend to be high or

TABLE 2

RATIO CHANGE FACTORS, 1954

Rank	County						1954 Ratio Change Factors
	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	
1						Sierra	1 17641
2						Alpine	1 11575
3						Mono	1 09705
4						Lassen	1 03972
5						Nevada	1 00876
6						Mariposa	97754
7						El Dorado	94910
8						Calaveras	94900
9						Plumas	93821
10						Amador	93343
11						Shasta	91791
12					Madera		91697
13						Siskiyou	.91636
14				Tehama			.90872
15					Tulare		.90690
16					Kern		90274
17						Imperial	89831
18			Riverside				.89829
19						Tuolumne	89258
20				Sutter			89226
21				Colusa			.89200
22			Ventura				88917
23						Lake	.88173
24						Inyo	.88071
25			San Bernardino				87895
26				Yuba			87793
27				Butte			.87707
28						Trinity	87673
29					Stanislaus		.87328
30					Merced		.87277
31						Modoc	87101
32					Kings		86704
33						Placer	86630
34						San Luis Obispo	85772
35				Glenn			.85288
36					Fresno		85051
37						Mendocino	.84859
38						Humboldt	84733
39						Santa Cruz	84710
40				Napa			84708
41			Santa Barbara				.83682
42			Orange				83305
43					San Joaquin		.82253
44	Los Angeles						82002
45						Monterey	81967
46						Del Norte	.81307
47	San Diego						81225
48				Yolo			79582
49		Sacramento					79505
50						Sonoma	.78785
51						San Benito	78562
52		Santa Clara					.78560
53		Contra Costa					77657
54	San Francisco						.75723
55	Alameda						.75502
56		San Mateo					.71872
57		Solano					.70025
58		Marin					.69210
						California	.81668

low in a certain year and the pattern testing, as shown on Table 2, helps identify such occurrences As stated earlier, there are no absolute standards for these factors, but they may be evaluated *relative to each other* within the group patterns, a county being known by the company it keeps. Further, this type of scanning serves the double purpose of defending the population estimates and providing confidence for them. For example, by itself, it might seem, that the ratio change factor, for San Mateo County (56th line) is too low, but material confidence is provided by noting the proximity of those of Santa Clara, Contra Costa, San Francisco, Alameda, Solano, and Marin Counties.

It is obvious that accurate ratio change factors require very accurate birth rates, which in turn require very accurate population estimates six to fourteen years earlier. It is considered possible to prepare intercensal estimates which are substantially accurate. In 1953, the California Taxpayers' Association re-estimated the population for each county for the intercensal years from 1940 to 1950 (also 1930 to 1940), using an adaptation of procedures outlined by Donald J Bogue [1] Estimated intercensal county populations as of January 1 were published [4]. Estimates as of each July 1 were also computed (and mimeographed but not published); new birth rates were calculated for each year from 1935 onward, and, new ratio change factors were produced from them.

Ratio change factors are for the most part rooted back in intercensal years for which data should be fairly sound By 1960, the years included will be 1946 to 1954, the last four of which will depend on four years of postcensal estimates of population. One, we trust, will not go too far wrong in four years, but any errors in population estimates of past years will be "inbred" into the change factors. Again, testing still to be discussed is designed to offset such errors.

CONVERSION FACTOR

The most difficult factor to determine is the "conversion factor." It is so called, as already stated, because it converts the ratio change factors into correction factors.

Necessity for the conversion factor is demonstrated on Table 3 using four years of presumably firm experience prior to the 1950 Census. Column 1 shows the school counts as of January 1 for the four years of 1947 to 1950 Column 2 shows the appropriate ratio change factor and column 3 shows the product of the school count times the Census ratio times the ratio change factor The products do not quite agree with the intercensal estimates of column 4 so we divide the intercensals by

the products to determine the factors by which the products must be multiplied to equal the intercensal estimates. These are shown in column 5 and prove to be a slowly ascending nearly straight line series.

Entering these factors on the aforementioned chart showing birth rates and ratio change factors, assists in visualizing the relationships. The birth rates (column 2 of Table 1) declined slowly in a roughly straight line to a low in 1933, then rose rapidly to a high in 1947, after which they again declined slowly to 1950. With the birth rates rising rapidly in this way, it seems not unreasonable that the ratio

TABLE 3

CALCULATION OF CONVERSION FACTOR, CALIFORNIA

Date	School Count	Ratio Change Factors	Product*	Intercensal Estimate	Conversion Factors†
	(1)	(2)	(3)	(4)	(5)
January 1947	1,020,218	1.15079	9,989,660	9,690,560	.97006
January 1948	1,086,317	1.09046	10,079,245	9,915,820	.98379
January 1949	1,165,313	1.03500	10,262,298	10,184,600	.99243
January 1950	1,237,263	.98885	10,410,082	10,505,896	1.00920

* School count times 1950 Census ratio of 8.50867 times ratio change factor
† Column 4 divided by Column 3

change factors derived from them would be affected in a manner analogous to the way arithmetic averages are affected by inclusion of extreme items. Thus, the inclusion of new years with very high rates causes overstatement of increase of density of children of grade school ages. Estimates produced with these ratio change factors are too low and a compensating adjustment is needed A similar correction in the other direction would become necessary should the birth rate turn rapidly down. Under nearly all conditions other than a stable birth rate, a correction will be needed. The conversion factor—one single statewide multiplier applied in common to all the county ratio change factors for a specific year to raise or lower all alike—is designed for this purpose.

Determination of conversion factors for years subsequent to the Census involves serveral different lines of attack all brought to bear on the one central problem.

1. *Extrapolation*

The four years shown on Table 3 serve as a base for projecting the conversion factors and the initial step is simply graphic extrapolation of the charted line. The projected values are used to produce estimates of population for the state as a whole year by year subsequent to the

Census and those estimates are checked to see if the conversion factors are tenable. Latest experience in what is called "the 1954 cycle," during which the estimates as of January 1, 1955, were produced together with revisions for previous years [2] showed that findings stated in Table 3 still hold good with minor variations to 1955. Conversion factors used in the "1954 cycle" were as follows:

	Year	Conversion Factors	Change in Year
Intercensal experience:	January 1947	.97006	
	January 1948	.98379	.01373
	January 1949	.99243	.00864
	January 1950	1.00920	.01677
Postcensal projection:	January 1951	1.015	.00580
	January 1952	1.025	.01000
	January 1953	1.030	.00500
	January 1954	1 040	.01000
	January 1955	1.045	.00500

The slight lowering from a strictly straight line projection is supported by collateral findings.

2. *Sum of Birth Rates*

The birth rates dropped and picked up in each successive year's sum of cumulated birth rates are reviewed and their effect on the conversion factor considered. On this score, a very slight dropping below the straight line projection appears reasonable.

3. *Evaluation of Growth Components*

Growth derives from natural increase and net migration. Statistics are available on births and deaths by place of residence but not on in- and out-migration. In order to show historical patterns of net migration, California Taxpayers' Association made a study of population, natural increase, and net migration for the state and each county for the 24 years from 1930 to 1954 [3].

It seems reasonable to hypothecate that the volume of net migration is responsive to general economic conditions in the nation at large. To affirm or refute the hypothesis, the net migration from the twenty-four year study was related to eleven of the statistical series identified by the National Bureau of Economic Research as "statistical indicators of cyclical revivals and recessions" [12]. The specific series used were Percent of Labor Force Unemployed (both National and State), Failure Liabilities, New Manufacturing Orders (Durables), Bank Debits Outside New York City, Industrial Production, Wholesale Prices

other than Farm Products and Foods, Gross National Product, Average Weekly Hours in Manufacturing, New Incorporations, Non-Agricultural Employment, and Freight Car Loadings These are selected from the "leading" and "concurring" groups of statistical series. Over the long run, the hypothesis that migration is correlated with economic activity appears supportable This conclusion, however, depends again not so much on "absolute" considerations of direct linkage as on "relative" considerations of turning at the same time in the same direction and being "less than" or "more than" a reference period.

Components of growth since 1940 as developed during the 1954 cycle are shown on Table 4 The natural increase has risen steadily except for 1948. The net migration has ranged from the wartime gain of some 570,000 in 1942 to the postwar low of about 76,000 in 1947. Gains during 1952 which were linked to the rearmament program for the Korean War almost equaled the 1942 gains but were not followed in 1953 by gains corresponding to those of 1943.

In determining the conversion factors for the 1954 cycle, the net migration they produced through 1953 was taken into consideration. After receipt of the October 1954 school count and completion of all calculations for estimates as of January 1, 1955, it was found that the produced net migration for 1954, shown on the first line of Table 4, was slightly greater than during 1953. At that time (December, 1954), the first half to three quarters of 1953 were available for the eleven "leading" or "concurring" economic indicators. The preponderance of evidence indicated a level of economic activity slightly below 1953 For consistency of our hypothesis, then, net migration in 1954 should be a little less than in 1953 rather than a little more As of the date of writing (January, 1955) there is evidence of economic revival during the latter part of 1954 Without elaborating, let us merely observe that when the record on 1954 is complete, slightly greater net migration may not be an inconsistency.

4. *Border Agricultural Inspections*

All autos entering California are stopped at the border agricultural inspection stations. Cars with non-California licenses and their passengers are tabulated Quarterly surveys determine reason for entering California such as "vacation," "moving to California," etc With these and other data on travel, it is possible to construct an approximation of in-migration to the state There is rough correlation with net migration but it takes some interpreting. Again, 1954 travel to the date of writing was a little below the 1953 volume.

TABLE 4

COMPONENTS OF POPULATION GROWTH FOR CALIFORNIA, 1940 TO 1954

Calendar Year	Components of Growth During Year		
	Total	Natural Increase	Net Migration
1954	455,400	E 197,105	258,295
1953	418,600	187,431	231,169
1952	712,900	172,898	540,002
1951	465,000	156,489	308,511
1950	276,070	145,874	130,196
1949	271,296	144,615	126,681
1948	268,780	141,093	127,687
1947	225,260	148,577	76,683
1946	231,100	122,866	108,234
1945	307,165	90,236	216,929
1944	406,200	88,321	317,879
1943	602,625	85,311	517,314
1942	640,370	69,716	570,654
1941	391,300	44,247	347,053
1940	243,272	32,545	210,727

E—Estimated.

5 *Comparison with Estimates by other Agencies*

Estimates for recent years are compared to estimates by other agencies, principally those by the United States Census Bureau [14, 15] and the California State Department of Finance [8]. Latest estimates of total California population, adjusted to July 1, (with percentage differences regardless of sign from the California Taxpayers' Association estimates in parentheses) are as follows:

	Calif. Tax Assoc	Bureau of the Census	Calif. Dept. of Finance
July 1, 1951	11,014,500	11,038,000 (0 2%)	11,115,000 (0 9%)
July 1, 1952	11,603,450	11,758,000 (1 3%)	11,612,000 (0.1%)
July 1, 1953	12,169,200	12,190,000 (0.2%)	12,075,000 (0.8%)
July 1, 1954	12,606,200	12,554,000* (0.4%)	12,450,000 (1 3%)

* Received after completion of California Taxpayers' Association estimates for January 1, 1955, but included for comparison

Such close agreement may be regarded as significant confirmation for all three series.

6. *Estimates of Population from Mortality Data*

Independent estimates of the total population of the state for recent years are made using age-specific deaths and death rates. This involves extension of death rate experience Estimates of the age cohorts up through age 14 are the most erratic and these are confirmed or supplanted by estimates of those cohorts from birth and school data. This in turn gives some current information on current death rate levels and the continuation (or non-continuation) of trend for these cohorts. Estimates by mortality vary from estimates based on school counts but do contribute to the over-all context for determining the conversion factors

7. *Grade to Grade Progression in Elementary School*

This line of study under stable conditions contributes to evaluation of gains by net migration but is currently complicated by the change of entry age for grade 1 already mentioned.

These and other lines of attack are continuing projects to help reduce the element of judgment in determination of the conversion factors and replace or support it with research considerations. Each year the components for the most recent years are revised and refined and from them as a base the conversion factor for the next year is projected.

Note that among other things the conversion factor compensates, first, for the half-year discrepancy between the date of the ratio change factors, which is July 1, and the date of the population estimates, which is January 1, and, second, for the error resulting from the use of annual totals of births in computing birth rates for the ratio change factors rather than their accumulation by months to agree exactly with the six to fourteen year age span at the Census date.

TESTING AND ADJUSTMENTS

Referring back to the formula, we are now down to the item "plus or minus adjustments" As already stated, one must always recognize that the ratio change factor for any county in any year may be in error because population estimates for one or more of the included years may have been wrong, or births may have occurred at an abnormally high or low rate in some years Can we, then, bring into play some independent element to test the accuracy of the estimates of population of individual counties as represented by the "raw product" of

school count times Census ratio times ratio change factor times conversion factor?

For this purpose a testing procedure is used involving four statistical series—place-of-residence births and deaths, auto registrations, and voter registrations. In form, the test is to see if a *change* in the relative portion that a county's population comprises of the state's population is matched by a change in its portion of the state's total of these four series. Again, any one of the four test series may be off-normal for a particular county in a particular year, so confirmation of the change for two consecutive years is required. For identification, this is referred to as "vertical series testing " The name derives from the columns showing each of the test series year by year converted to percentages of the state total

This testing is most easily explained by illustration. In April, 1950 Los Angeles County had 39 218 per cent of the state's total population, and in January, 1954, according to the unadjusted estimate (the "raw product"), it had 40.856 per cent representing *an increase* of 1.638 percentage points. Since we have estimated that Los Angeles had a larger portion of the state's population in 1954 than in 1950, it is reasonable to assume it should have produced a correspondingly larger portion of the state's births In 1950 Los Angeles had 36 411 per cent of the state's births and in 1954, 37.170 per cent, *an increase* of .759 percentage points. Thus, we find that Los Angeles is, in fact, credited by the State Bureau of Vital Statistics with having had a larger portion of the births, but the reported gain in the county's portion of births, .759 percentage points, is not as great as the estimated gain in portion of the population, 1.638 percentage points. Does this mean that the population estimate as represented by the "raw product" is too high? Perhaps, but before accepting this interpretation, we should see if it is confirmed by the other series similarly treated over two or more consecutive years The differences in percentage of total from 1950 to the indicated years were as follows:

Los Angeles County	1953	1954
Population—calculated	1.079	1.638
—adjusted	.679	1.238
Births	.759	1.322
Deaths	.679	−1.452
Autos	.760	1.278
Voters	Not Available	.075

The positive differences in Los Angeles's portion of the total state population of 1.079 in 1953 and 1 638 in 1954 were greater than the differences in all of the seven test items, hence, the larger fraction of the state's population is not supported by correspondingly larger fractions of the total produced by the county in each of the four series. In other words, the test indicates that the "raw product" estimates of the population of Los Angeles County for the two years are too high Reducing the calculated population of Los Angeles County by 0 4 per cent of the state total population for both years gives the "adjusted" differences shown above. The adjusted differences are *approximately central among the test differences* If we, then, adjust the estimate for Los Angeles County by deducting 0 4 per cent of the state total, which was about 50,000 in 1954, the gain in portion of population will be more closely supported by the gain in portion of births, deaths, voters, and autos actually tallied for the county

In the "1954 Cycle" adjustments up or down were applied to the "raw products" of 22 counties They ranged from .4 points in Los Angeles County down to 01 points in a few of the smaller counties. The total of the adjusted estimates for the counties, after rounding, was taken as the estimate of population for the state Where no clear indication of need for an adjustment occurs, it is assumed the population figure for a county is about right or decision is suspended until the next annual cycle.

It will be recognized that the vertical series test procedure is different from plain pro-rating. As pointed out in the discussion of the Census ratio, Madera County has a ratio less than half that of San Francisco because its birth rate is so much greater On a simple pro-rata basis using births, Madera County would be given too large an estimate of population and San Francisco too small an estimate Every county has its own peculiar character and its own specific production of births, deaths, voters, and auto registrations Using the differences from the vertical series avoids the biases of pro-rating, but involves the assumption that each county's own peculiar character continues from year to year. This assumption is weakest with respect to voter registrations since local issues may affect registrations rather markedly

Part of the process of determining adjustments from the vertical series testing is to refer back to the pattern testing of the ratio change factors. If a "raw product" appears high according to the vertical series test, a check is made to see if the corresponding ratio change factor appears high on its scale Vertical series adjustments may be made without these confirmations, but in many instances there is confirma-

tion. Checking out the vertical series tests for all the counties and relating them to the pattern tests is laborious, but it is believed that the resulting adjustments are *in the right direction* and consequently reduce margins of error Thus, the elements used to estimate each county's population are not evaluated singly in isolation but only in context with those of all the other counties. Through these tests, changes in demographic or economic character can be detected.

COVERAGE

Estimates produced by this method are of total resident population. No effort is made to deduct military personnel because data on military personnel are usually classified and current estimates of total population fulfill many uses without separating the military Using the procedures here outlined, it is obvious that it would be necessary to restrict grade school counts to the children of the civilian population alone in order to produce estimates of civilian population That, of course, is impossible.

The 1950 Census counted individuals on military and naval posts where assigned. This included both individuals permanently assigned to the bases such as administrators, instructors, operators, etc , as well as those temporarily assigned, such as trainees being organized into units Hence, the 1950 Census ratios included civilians together with permanently assigned military personnel and a *corresponding load* of temporarily assigned. During times of rapid military expansion, as in 1951 and 1952, permanently assigned personnel may carry relatively larger loads of temporarily assigned than in April, 1950, when military establishments were at a low ebb making the 1950 Census ratios to the extent of the extra load not representative.

Californians temporarily assigned in a county in the state other than that of residence, or assigned outside the state on military duty are included in estimates made by this procedure because their families remain in the county of residence and are reflected in the school counts. Californians assigned within the state but outside their county of residence may offset one another somewhat. Also Californians assigned outside the state and non-Californians assigned in the state may offset one another The net errors are probably small in proportion to the total population of most counties It could be appreciable in some, however. The error, when there is one, is probably that of not fully reflecting temporarily assigned military personnel. This does not preclude use of the estimates for many practical purposes and while the military "float" may mean business for local amusements, personal

services, and retailers, it is perhaps better discounted for evaluating real growth and for many research purposes.

TESTS OF ACCURACY

Comparison of estimates of population as of April 1, 1950 with the 1950 census for the twenty-two larger California counties is shown in Table 5 [5] California has 58 counties, of which the first 22, containing 91 per cent of the state's population, are arbitrarily regarded as "larger" because of their general character.

Of the 22 "larger" counties, 14 had errors under 5 per cent, leaving eight with errors over 5 per cent, of which one, Monterey, had an error exceeding 10 per cent. The error of estimate for the state as a whole was 0.9 per cent and error of estimate for the 22 counties as a whole was 0.3 per cent. Among the 36 "smaller" counties, 20 of which had fewer than 20,000 inhabitants in 1950, 11 had errors less than 5 per cent. Of the 25 with errors exceeding 5 per cent, 15 had errors exceeding 10 per cent and 4 exceeding 25 per cent.

In order to compare the estimates with some "bench mark" measures of accuracy in population estimating, we may compare the percentage deviations with percentage deviations of estimates for states and cities made for test purposes by the Bureau of the Census using various techniques [10] Comparisons are shown in Table 6. Average deviations for the "larger" counties are slightly greater than those of the Census Bureau for states using "Migration and Natural Increase-Method II." Deviations of the "smaller" counties and of all 58 counties as a group are greater than those reported for states by the Census Bureau for any method.

It is important to know in this connection that the basic statistical series—count of enrolment in grades 1 through 8—was changed in 1947 from the former statistically inferior series known as "State Enrolment" to the current statistically excellent series of counts of active enrollments on each March 31 and October 31 As a result of this mid-decade change, it was necessary to use estimates based on correlations between the old and the new school counts in setting the 1940 Census ratios. Tests—which need not here be explained in detail—indicate that the large errors among the "smaller" counties were due substantially to errors introduced by the 1947 change in counting method. For the current decade we have the improved school counts together with improvement in others of the materials used for population estimating. It is not unreasonable to assume, therefore, that results in 1960 will be as good as and probably better than they were

TABLE 5

COMPARISON OF POPULATION ESTIMATES OF APRIL, 1950
BY CALIFORNIA TAXPAYERS' ASSOCIATION
WITH 1950 CENSUS FIGURES

County	Census, April 1, 1950		C.T.A. Estimate, April 1, 1950	Deviation of Estimate from Census	
	Population	Per Cent of State Total (Cumulative)		Persons	Per Cent
Los Angeles	4,151,687	39.22	4,141,600	−10,087	−0.2
San Francisco	775,357	46.54	778,600	3,243	0.4
Alameda	740,315	53.53	761,200	20,885	2.8
San Diego	556,808	58.79	567,300	10,492	1.9
Contra Costa	298,984	61.61	300,300	1,316	0.4
Santa Clara	290,547	64 35	288,100	−2,447	−0.8
San Bernardino	281,642	67 01	287,500	5,858	2.1
Sacramento	277,140	69 63	289,500	12,360	4.5
Fresno	276,515	72 24	290,900	14,385	5 2
San Mateo	235,659	74 47	233,700	−1,959	−0 8
Kern	228,309	76.63	223,100	− 5,209	− 2.3
Orange	216,224	78.67	201,200	−15,024	− 6.9
San Joaquin	200,750	80.57	204,900	4,150	2.1
Riverside	170,046	82.18	166,000	− 4,046	− 2.4
Tulare	149,264	83.59	159,900	10,636	7.1
Monterey	130,498	84.82	109,600	−20,898	−16.0
Stanislaus	127,231	86 02	131,500	4,269	3 4
Ventura	114,647	87 10	103,500	−11,147	− 9.7
Solano	104,833	88 09	110,100	5,267	5.0
Sonoma	103,405	89 07	108,500	5,095	4.9
Santa Barbara	98,220	90.00	89,800	− 8,420	− 8 6
Marin	85,619	90 81	92,400	6,781	7 9
Sub-Total	9,613,700	90 81	9,639,200	25,500	0 3
Rest of State	972,523	9.19	1,037,030	64,507	6.6
STATE TOTAL	10,586,223	100.00	10,676,230	90,007	0.9

TABLE 6

COMPARISON OF PERCENTAGE DEVIATIONS OF ESTIMATES OF CALIFORNIA COUNTIES BY CALIFORNIA TAXPAYERS' ASSOCIATION METHOD FROM THE 1950 CENSUS WITH DEVIATIONS OF ESTIMATES FOR STATES AND CITIES BY SELECTED METHODS

	Average Deviation	Quadratic Mean Deviation	Deviations of 10% or More	Deviations of 5% or More	Positive Deviations
California Taxpayers' Ass'n. Method:					
California Counties					
"larger" (22)	4.37	5.74	1	8	13
"smaller" (36)	11.51	15.79	15	25	27
all (58)	8.79	12.93	16	33	40
*Other Methods:**					
States (49) (Adjusted to national total)					
Migration and Natural Increase					
Method I	5.84	7.38	10	22	
Method II	3.47	4.78	2	11	
Vital Rates	4.38	5.54	4	19	
Arithmetic	6.39	8.01	11	24	
Geometric	6.35	8.01	11	23	
Cities (92)					
Migration and Natural Increase					
Method I	8.34	9.86	29	65	30
Method II	6.53	8.52	18	50	22
Vital Rates	9.33	12.56	31	59	79
Arithmetic	9.60	12.15	33	65	11
Geometric	9.33	11.75	31	64	13

* Source. See reference note 10.

in 1950. There is no way of judging what degree of improvement for both "larger" and "smaller" counties will result from the better components of this decade but it appears, both on the basis of rationale and possible results, that the procedure warrants continued use and development.

REFERENCES

[1] Bogue, Donald J., "A technique for making extensive population estimates," *Journal of the American Statistical Association*, 45 (1950), 149–63.

[2] Brown, Hugh H., "Population of California counties—Estimates as of January 1st," *Tax Digest*, California Taxpayers' Association, Vol. 32 (December, 1954) 416–7.

[3] Brown, Hugh H., "People come—and people go," *Tax Digest*, California Taxpayers' Association, Vol. 32 (March, 1954), 92–8.

[4] Brown, Hugh H., "Intercensal estimates of population, January 1, for years between 1940 and 1950 Censuses," *Tax Digest*, California Taxpayers' Association, Vol 32 (January, 1954), 9–10.

[5] Brown, Hugh H., "Population of California counties—Final 1950 Census data released," *Tax Digest*, California Taxpayers' Association, Vol. 29 (June, 1951), 188–9, 212–3.

[6] Eldridge, Hope T., "Suggested procedures for estimating the current population of counties," *Population-Special Reports*, U. S. Bureau of the Census, Series P-47, No 4.

[7] Frisén, Carl M., "A technique for estimating population from mortality data," *unpublished doctoral dissertation*, Graduate School of Business, Stanford University, January, 1951

[8] Frisén, Carl M., "Estimated population of California, 1950–1954, with projections to 1965," *California Department of Finance release*, July, 1954, Table 1

[9] Greenberg, Benjamin, and Lawrence, Norman, "Illustrative examples of two methods of estimating the current population of small areas," *Current Population Reports*, U. S Bureau of the Census, Series P-25, No. 20

[10] Greenberg, Benjamin, Siegel, Jacob S., and Shryock, Henry S., Jr., "Accuracy of postcensal estimates of population for States and cities," *American Sociological Review*, 19 (1954), 440–6.

[11] Lawrence, Norman, and Shryock, Henry S., Jr, "The current status of State and local population estimates in the Census Bureau," *Journal of the American Statistical Association*, 44 (1949), 157–73.

[12] Moore, Geoffrey H, *Statistical Indicators of Cyclical Revivals and Recessions*, National Bureau of Economic Research, Inc., New York. Occasional Paper 31.

[13] Schmitt, Robert C., "Short-cut methods of estimating county population," *Journal of the American Statistical Association*, 47 (1952), 232–8.

[14] U S Bureau of the Census, "Estimates of the population of continental United States, by regions, divisions, and States, and of Alaska, Hawaii, Puerto Rico, the Canal Zone and the Virgin Islands: July 1, 1950 to 1953," *Current Population Reports*, Series P-25, No 97.

[15] U S Bureau of the Census, "Provisional estimates of the population of states, July 1, 1954," *Current Population Reports*, Series P-25, No. 108.

THE MEMORY FACTOR IN SOCIAL SURVEYS

PERCY G. GRAY, *Government Social Survey, London*

This article considers the problems of framing questions, the answers to which depend on the informants' memories. The problems are illustrated with examples from the British Survey of Sickness. In an experiment in which a group of government employees were asked to recall what sick leave they had taken and when, the predominant error consisted of forgetting *when* the leave occurred rather than forgetting it completely. The results of checks on other memory questions show the importance of carefully defining and redefining for the informant the period of time a question covers.

1. THE PROBLEM

TO OBTAIN accurate answers to many of our questions in social surveys we have to depend very much upon the memories of our informants In framing such questions two queries immediately arise. Firstly, should we ask about a period fixed in time, e g , a particular week, or should we ask about a period ending on the day before the interview, e.g , the last seven days? Secondly, how long shall the period be, a day, a week, a month, or what?

The first query arises in this way No matter how much one may wish to do so, one cannot interview all of a randomly selected sample of people on a given day An appreciable proportion of people will not be at home when the interviewer calls. Furthermore, practical considerations usually dictate that the interviewing shall be spread over at least a week, probably longer.

Thus if we choose the fixed period of, say, the week January 1st–7th then, if the interviewing is spread over a fortnight, the maximum period of recall demanded of informants will vary from 7 days up to 21 days. (A simple analysis by date of interview will not necessarily show whether this has had any effect, since the groups interviewed on different days are not comparable. With a sample of adults, for example, higher proportions of housewives are likely to be interviewed in the early days of the interviewing period.)

If, however, we ask about the seven days prior to the interview there are two difficulties. The most obvious is that different people will be asked about different periods of seven days. In many cases, there is little objection to this and one is prepared to treat the results as applying to an average week. The second difficulty, which should not, however, be exaggerated, is the possibility that there is some relation-

ship between the informant being found at home and the variable that one is studying. This may be troublesome when the memory period is reduced to "yesterday," i.e., the day before the interview. Consider an individual who is out on alternate evenings. If the interviewer's first call is made on an evening when the informant is out, she may well make an appointment for the following night when the informant would be in. The effect of this might be to over-estimate, say, evening expenditure on entertainment when based on "yesterday."

These are the difficulties which face one when deciding whether to ask about a period of constant length which is also fixed in time, or a period of constant length but ending on the day before the interview. The next point we will consider is the length of period to be adopted.

Let us suppose we are dealing with medical consultations and that we wish to calculate the average consultation rate for a four-weekly period. In the first place let us ignore some of the practical interviewing difficulties which we have already mentioned, and assume that we can interview as many people as we like on any day we care to choose. Then we might make all our interviews on one day and ask about the past four weeks. Or at the end of the first week we might question a quarter of the total sample about that week and question a second quarter a week later about the second week, and so on. Or we might subdivide our sample into 28 parts, spread the interviewing over 28 days, and ask only about the day before the interview.

At first sight it might appear that the standard error of the estimated four-weekly rate would vary as $1:\sqrt{4}:\sqrt{28}$ for the three cases. But this is not so: although we obtain information about four times as many weekly periods in the first case as in the second, it is the same group of people who are reporting about four consecutive weeks, and there is a tendency for the same people to see the doctor in each week.

This point can be illustrated by some data which are available for rather different periods. In Table 1 are shown the number of consultations made at doctors' offices by a sample of 3533 adults in the two consecutive calendar months of December and January. The first column gives a distribution of the consultations for December, the second column for January, and the third for the two months combined.

Now if the same random sample of 1,000 persons is asked about their consultations in each of the two months, the standard error of the average for the two-monthly period will be $1.300/\sqrt{1{,}000}=0\ 0411$. If, however, one sample of 1,000 persons is asked about December and another separate sample of 1,000 persons is asked about January, then

the error of the estimated *two*-monthly average is

$$[\{(0.731)^2 + (0.811)^2\}/1{,}000]^{1/2} = 0.0345.$$

The practical decision one has to make is as follows If one can only afford to make 1,000 interviews, should 500 be made at the beginning of January dealing only with December, and a further 500 at the beginning of February dealing only with January, or should all the 1,000 interviews be made at the beginning of February and deal with both of the two months? The sampling error for the estimated two-monthly average with the first procedure as compared with the second, is increased by the factor $(0.0345/0\ 0411)\sqrt{2}$, not $\sqrt{2}$ as would be approxi-

TABLE 1

PERSONS HAVING DIFFERENT NUMBERS OF CONSULTATIONS AT A DOCTOR'S OFFICE IN TWO CONSECUTIVE MONTHS

Number of consultations	December	January	December and January
0	3,054	3,015	2,763
1	277	298	348
2	109	117	190
3	43	47	82
4	41	41	78
5	6	4	28
6	1	6	12
7	—	2	5
8	1	1	19
9	1	1	4
10		—	1
11		—	—
12		—	—
13		1	1
14			—
15			—
16			1
17			—
18			1
Total number of persons	3,533	3,533	3,533
Average number of consultations per person in the period	0.238	0.265	0.533
Standard deviation	0.731	0.811	1 300
	σ_1	σ_2	σ_{1+2}

For the two months the correlation coefficient $\rho_{12} = +0\ 42$.

The three standard deviations and the correlation coefficient are of course connected by the relation—$\sigma_{1+2}^2 = \sigma_1^2 + \sigma_2^2 + 2\rho_{12}\sigma_2\sigma_1$.

mately the case if there were no correlation. It is against this somewhat smaller increase that we have to weigh the possibly greater error due to memory, which may arise through asking people to remember their experience for two months instead of one.

With many variables dealt with in survey work there is likely to be this positive correlation over consecutive periods of time. The correlation may, however, be quite different between, say, consecutive days and between consecutive weeks. Occasionally there may be a negative correlation such as is found in the example of Table 2

In this case the standard error of the two-monthly average which

TABLE 2

ANNUAL LEAVE TAKEN BY 461 PERSONS IN TWO CONSECUTIVE MONTHS

	July	August	July and August
Average number of days per person	3 72	4.30	8.02
Standard deviation	4 64	4 69	5.70
	σ_1	σ_2	σ_{1+2}

For the two months the correlation coefficient $\rho_{12} = -0\ 25$.

would be obtained from two separate samples of 1,000, each asked about one month only, would be greater, not less, than that for one sample of 1,000 asked about the two months

Thus far we have considered the gains likely to result from increasing the length of the memory period if there were no memory losses. But these gains have to be weighed against any loss in accuracy due to memory. Such loss can take two forms. In the first place, the variance of the remembered values may be greater than the variance of the true values. In the second place, the average of the remembered values may be biased. Even if there is no bias and the errors are fully compensating the variance may still be increased. The relation between the variance of the remembered values and that of the true values is as follows:

$$\sigma_r^2 = \sigma_t^2 + \sigma_d^2 + 2\rho_{td}\sigma_t\sigma_d$$

where

σ_r = standard deviation of the remembered values

σ_t = standard deviation of the true values

σ_d = standard deviation of the discrepancies (the discrepancy = remembered value—true value)

and

ρ_{dt} = correlation between the discrepancies and the true values.

We see that the variance of the remembered values will exceed the variance of the true values, unless the correlation between the discrepancies and the true values (ρ_{dt}) is negative and greater in value than $\sigma_d/2\sigma_t$.

It is possible in the normal pilot inquiry to determine approximately the reduction in the standard error to be expected from increasing the length of memory period, but it is generally impossible to measure the error due to memory. Indeed, we have found it extremely difficult to obtain information about memory errors even on the main inquiries. Sometimes it is possible to compare totals or averages with national estimates that are available, but the results are seldom exactly comparable and serve only to detect large errors. Asking questions about different periods is of limited value, since a comparison of the results obtained usually involves the assumption that the shortest period produces the smallest memory error. What is required is some experiments where the true and remembered values are known for each person questioned and such experiments are extremely difficult to arrange. In an attempt to learn something of the errors arising due to memory we did, however, devise a small experiment in November, 1951. But before describing the experiment something must be said of the Survey of Sickness, since it was experience with this survey which determined the lines of the experiment.

2. THE BRITISH SURVEY OF SICKNESS

The Survey of Sickness commenced in 1944 and continued with a few gaps until the beginning of 1952. Sampling and interviewing were carried out by the Social Survey, while the General Register Office dealt with the classification of the illnesses and analysis of the results. During the first fortnight of each month a different sample of adults was questioned about their illnesses in previous months. Until September, 1949, informants were asked about illnesses in the three calendar months prior to the interview. From September 1949 onwards they were only asked about the two previous calendar months

Thus, prior to September 1949 the sample interviewed in the first fortnight of, say, January was asked:

"Did you have any illness, ____________ during October, November, and December?"

The interviewer then went on to find out for each illness or injury,

whether it was present, and how many days of incapacity and medical consultations it caused, in each of the three months separately. From the data which accumulated month by month it was possible to calculate monthly, quarterly, and annual rates. Clearly if N persons are interviewed each month and there is no memory error, then the monthly rates can be based on the experience of $3N$ different people and the standard error will be reduced to $1/\sqrt{3}$ times that which would be obtained if informants had been asked about one month only. In calculating annual rates, however, it has to be remembered that although there are $36N$ monthly experiences, these are not the experiences of $36N$ different people. The equivalent size of sample is rather less than $36N$, due to the positive correlation existing between peoples' experiences in the three consecutive months about which they are asked.

As soon as the results became available for three consecutive samples, it was found that there were appreciable differences between the results obtained at different times after the month of experience. Some results taken from an early report [1] are given in Table 3.

It will be seen that there is no difference for December in the number

TABLE 3

THE EFFECT OF DIFFERENT INTERVALS BETWEEN THE MONTH OF SICKNESS EXPERIENCE AND THE MONTH OF INTERVIEW

Type of illness	Interviews conducted in	Monthly incidence per 100 at risk				
		Month of Sickness Experience				
		Oct.	Nov.	Dec.	Jan.	Feb.
Serious illness	January	1 7	2.1	2.0		
	February		1 8	2 0	1 8	
	March			2 0	2 0	2.1
Influenza	January	1.5	8.5	14.2		
	February		5.2	10.7	5.0	
	March			8 1	5 3	2.3
Colds	January	4.9	9.8	19.3		
	February		8.3	14.0	12.6	
	March			11.3	13.7	15.8

Note: The sample sizes were: Jan. 1,944, Feb 2 399, Mar. 2,402

of persons with a serious illness but that there are appreciable differences for influenza and colds These findings were confirmed by later results and in giving the results for January-April, 1945, it was stated [2] that "the minor ailment rates . . . are based only on the two months preceding the interviews since the recollection of minor complaints three months ago is not reliable."

Beginning with the March quarter of 1947 onward the results of the inquiry were presented in a series of tables of more or less standard form in the Registrar General's Quarterly Returns. With the exception of the first table giving certain overall rates, only data based on the two most recent months prior to interview were used This one table continued to be based on three months' experience until the end of 1948, when it was replaced by data relating to only the two months prior to interview. However, informants were still asked about their sickness experience in the three months prior to interview until the interviews made in September 1949 In this month the questionnaire and interview were modified so that informants were asked about only the two, not three, calendar months prior to the interview. This continued until the survey was suspended in March 1952. Plans were then afoot to reduce the memory period to one month only. This was to be done at first on half the sample only, in order to study the effect, if any, on the results for the most recent month. (Such a split sample was unfortunately not used in September 1949, when the reduction from the three month to the two month memory period was made.)

During this time the only published comments at any length made on the effect of memory are by Stocks in his report "Sickness in the Population of England and Wales in 1944–1947" [3] He concluded that "no advantage would accrue by basing morbidity rates on the last month's experience alone . even if the number of people interviewed each month was so increased that the sampling error would be the same " The argument[1] leading to this conclusion is rather involved and will not be reproduced here but subsequent examination showed that it was not soundly based. There was in fact no evidence in favor of a memory period of two months.

As Stocks realized, it may be very difficult to say when an illness started; a person's idea of the starting date may well change, quite apart from memory, as the illness progresses A more definite point in time would seem to be the first medical consultation and it is medical

---

[1] The analyses presented in reference [3] are based on only the most serious illness of each person not all illnesses as they should be

consultations which we shall consider. In Fig. 1 we give the variation in the number of medical consultations per person per month as reported at three different intervals after the month of experience. The top set of curves give the results for 1947 and 1948 according to interviews made in the month following the month of experience, and also from interviews made one and two months later. The bottom set for 1950 and 1951 consists of only two curves, as by then informants were only asked about two months, not three (1949 has been omitted since the results for the third month are not available from October 1948, onwards while in September 1949, the memory period was reduced to two months) It is at once apparent that the memory effect is not simply one of forgetting an increasing proportion of consultations as time goes by. It seems that the difference between the curves is affected by the way in which the actual rate is varying For example, it appears that when the actual rate is falling sharply from a high figure, e.g, during the early parts of 1947 and 1951, there is a tendency for the curves to cross.

This was the state of our knowledge in 1951 We suspected that most of the trouble arose because people were forgetting when consultations occurred rather than forgetting them completely. Ideally we would have liked to check each person's answers with, say, his doctor's records. Unfortunately, doctors do not in general keep sufficiently detailed records We had therefore to fall back on the records of sickness absence and annual leave kept by a Government department. The funds and time available were a further serious limitation on what could be done.

3. AN EXPERIMENT

We chose for our experiment a group of Government offices where the records of annual leave and sick leave taken by the employees could be made available to us Our investigators were supplied with forms similar to that shown in Figure 2 and a list of room numbers. On entering a room the investigator explained the purpose of the inquiry, handed out a form to each person present, and gave brief instructions. As the investigator collected the forms she asked the people present not to talk to anyone else about the experiment until lunchtime On leaving the room the investigator added the room number to her list of rooms visited so that we could tell the order in which they had been done. All the forms were completed during a period of about two hours on the morning of November 15th.

Our choice of the 15th of the month was made to simulate the worst

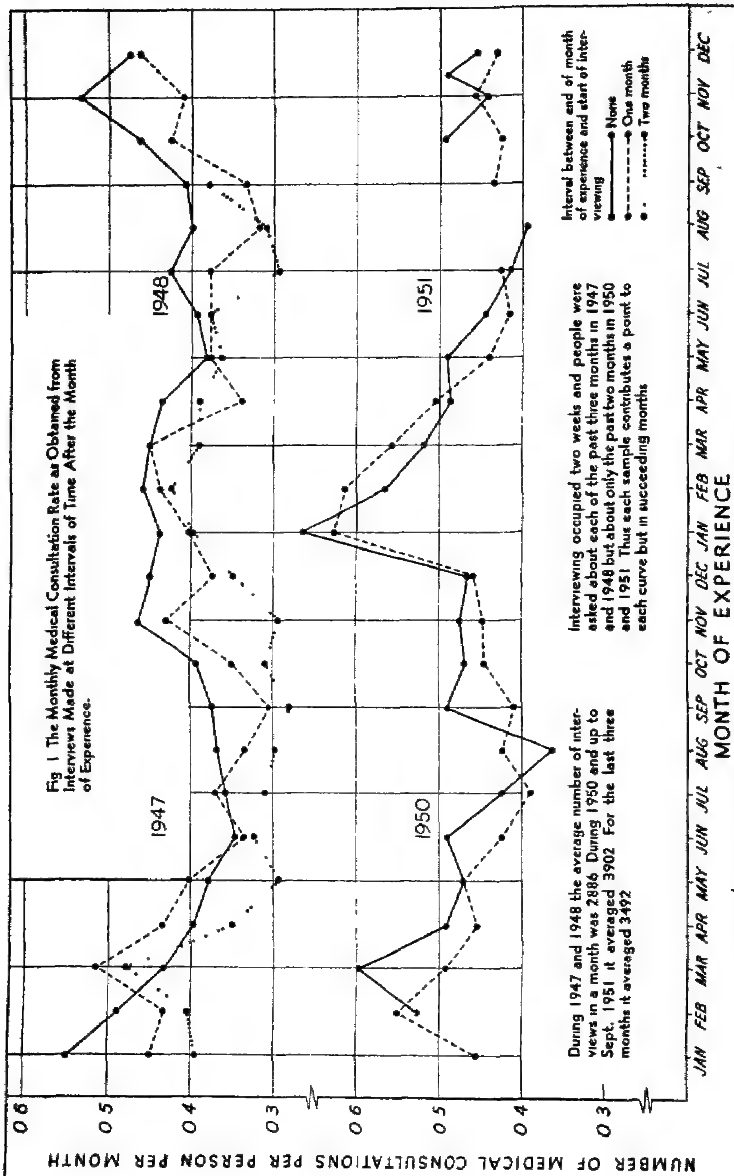


Fig. 1. The Monthly Medical Consultation Rate as Obtained from Interviews Made at Different Intervals of Time After the Month of Experience.

condition in the Survey of Sickness interviewing where interviews were normally spread over the first fortnight of each month. We would have liked to have varied the memory period covered, i e., the number of months, but with the limited time and resources at our disposal we could not afford to split the sample. We settled on the four calendar months of July, August, September, and October, and the part of November as the period we would ask about. In getting people to fill in the forms we were, of course, departing from an interview proper, but this was the only way to get appreciable numbers of cases with the time and labor available. It seems doubtful however that an interview would have produced better results.

Although we asked people not to talk about the inquiry, a few

PLEASE DO THIS FROM MEMORY

EVEN IF YOU DON'T THINK YOU CAN REMEMBER CORRECTLY, WOULD YOU MAKE THE BEST ESTIMATE YOU CAN

ANNUAL LEAVE

MONTH	*ANNUAL LEAVE* (Number of days *taken*)
JULY	
AUGUST	
SEPTEMBER	
OCTOBER	
NOVEMBER (to date)	

If you had *no* leave in any month write in "NONE"

* * *

If you had more than one period of leave within a month, please enter as follows 1+2, or 6+3, etc

* * *

SICK LEAVE

MONTH	*SICK LEAVE* (Number of days) *taken*
JULY	
AUGUST	
SEPTEMBER	
OCTOBER	
NOVEMBER (to date)	

LEAVE THIS BLANK

Room Number ____________________

Name (Block Capitals) ____________________

Division ____________________

FIG. 2. Form used in the experiment

certainly told their friends in other offices. In the few cases where people were known to have heard of the inquiry in advance, their forms were discarded. It is probable, however, that a small number of those who got completely correct forms had looked up the answers. The effect cannot be very great since there were so few all-correct forms. Furthermore, when the forms were divided into three groups according to how early they were completed, we find that there was no "improvement" in memory as the morning wore on

The sick leave and annual leave taken by each person was later obtained from the department's leave records and added to the forms. In addition to the period covered by the forms the leave taken in the earlier month of June was also recorded There is no reason to doubt the accuracy of the leave records and, in what follows, it has been assumed that they provide the true values of the leave taken. During this process of adding the true values to the forms a few were found to be incomplete and a few were found where the informant had had special leave which he might, or might not, have counted These were discarded The remaining 433 forms provide the analyses which follow.

In the case of annual leave there is an unusual factor in that each person is entitled to a certain definite number of days, varying with grade, during the twelve months commencing February 1st Thus although a person might not know just when he had taken leave he might have a good idea of the total amount taken Sick leave[2] does not have this disadvantage and will be dealt with first.

Of the 433 people there were 205 who, according to the staff records, had not had any sick leave during the months in question Of these 192 recorded none on their forms Of the remaining 13 there were 11 who gave a figure for one month and 2 who gave figures for two months

There were 144 persons who had taken sick leave in one month only Of these 59 recorded both the amount and month correctly on their forms, 25 gave the correct amount but the wrong month, while 22 got the month right but the amount wrong. There were 24 cases where no leave was recorded, and there were 14 cases where leave was recorded in more than one month though it had only been taken in one month.

Sick leave had been taken in two months by 63 people and of these 11 got both periods correct, 21 got one period correct, while the remaining 31 got neither period correct.

The remaining 21 persons had taken sick leave in three or more

[2] The detailed results can be obtained from the author.

months and of these 4 got all periods correct. The results for all 433 persons may be summarized in that of the 205 who had taken no leave, 192 gave the correct answers, while of the 228 who had taken some, only 74 gave completely correct answers. In view of this it is perhaps surprising to find that the average of the remembered values for the whole period differs from the true value by only 1 per cent A comparison between the remembered and true values for each month and for all the months together is given in Table 4

TABLE 4

AVERAGES AND VARIANCES FOR SICK LEAVE

	July	August	September	October	November 1–14th	All months
Average (days)						
True values	0 261	0 446	0 695	0 963	0 279	**2.644**
Remembered	0 275	0 415	0 724	0 855	0 341	2 609
Discrepancies	0 014	−0 031	0 029	−0 108	0 061	−0 035
Average discrepancy as a percentage of the average true value	5%	−7%	4%	−11%	22%	−1%
Variances						
True values σ_t^2	1 375	3 993	8 337	6 659	1 361	41 888
Remembered σ_r^2	2 744	4 642	9 536	6 202	1.763	45.219
Discrepancies σ_d^2	1 612	1 013	1 510	1 354	0 261	2 755
Ratio of the variance of the remembered values to the variance of the true values	1.99	1 16	1 14	0 93	1 29	**1.08**
Standard error of discrepancies	0 061	0 048	0 059	0 056	0 025	0 080
Correlation between discrepancies and true values ρ_{dt}	−0 082	−0 091	−0 044	−0 302	0 119	0 027

Discrepancy = Remembered Value − True Value.

When the average discrepancy is expressed as a percentage of the average true value, this varies from +22 per cent for the fortnight of November to −11 per cent for October. For all but one month the variance of the remembered values exceeds the variance of the true values. In this month there is, of course, an appreciable negative correlation between the discrepancies and the true values. In no case does the average discrepancy exceed twice its standard error Thus we cannot say that the average discrepancies differ significantly from zero, i e., that there is a residual bias Clearly an experiment on a much larger scale was required to determine this. Something can, however

TABLE 5

TYPES OF ERROR MADE IN REMEMBERING SICK LEAVE

	July	August	September	October	November 1–14th
Correct					
No leave taken	380	366	341	283	369
Some taken	13	9	28	56	26
	393	375	369	339	395
Transference					
In from later month	3	5	13	6	0
In from earlier month	4	2	11	17	21
May be in from earlier month, may be invention	6	8	2	8	4
	13	15	26	31	25
Transference					
Out to later month	4	12	12	19	0
Out to earlier month	0	1	4	12	6
May be out to earlier month, may be forgotten	20	16	13	11	4
	24	29	29	42	10
Wrong amount					
Overestimate	1	7	4	13	0
Underestimate	2	7	5	8	3
	3	14	9	21	3
All Persons	433	433	433	433	433

be learned of the types of error that were made. This analysis is presented in Table 5.

Each month has been considered separately. Thus for September there were 369 correct entries, 341 being cases where no leave was taken. There were 13 cases where leave, which had been taken in either October or November, was entered for September Then there were 11 cases where leave was recorded for September which was actually taken in August, July, or June. There were 2 cases where the leave was either purely imaginary or had actually been taken before June (we had no information about earlier leave).

On the other hand, there were 12 cases were leave actually taken in September had been recorded for either October or November. Similarly, there are 4 cases where it had been recorded for either August or July. Then there were 13 cases where the leave may have been completely forgotten, or where it may have been thought to have occurred prior to July. (We do not know what the informant thought about earlier months).

Finally, there were 4 cases where the informant remembered some leave for September but overestimated it, while in 5 cases the leave was remembered but underestimated.

In making this classification a period of leave, which was remembered wrongly both as to amount and month, has been classified as a transference Also any period wrongly recorded as "none" appears as a transference not an underestimate It will be clear that in making such a classification a certain amount of judgment has been used and the results must only be taken as indicating the order of things. In comparing months it must be remembered that we had not available the actual leave taken prior to June, and we do not know what the informant thought he had taken prior to July.

In spite of these limitations, the classification is of value in that it draws attention to the importance of transference and the compensating nature of many of the errors. Forgetting mainly took the form of forgetting *when* rather than forgetting completely. With so much transference occurring, there is every reason to expect that estimates for a particular period of time will be affected by experience prior to this period, as would appear to be the case with medical consultations in the Survey of Sickness. There would appear to be a case for taking the memory period up to the day before the interview, if this is possible, since this closes the door to one set of transfers.

Analysis of the data on annual leave gave similar results to those for sick leave. The averages and variances are given in Table 6.

Again, as in the case of sick leave, the variance of the remembered values exceeds the variance of the true values, except for the month of October. Again, the small sample size makes it impossible to say whether the average discrepancies differ significantly from zero.

An attempt to analyze the types of error, as in Table 5, proved more difficult than in the case of sick leave as the number of periods of annual leave taken by each person was much greater. Nevertheless it was again clear that transference was the major factor

This experiment left us with many questions unanswered. What would be the effect of not asking about the fortnight in November?

TABLE 6

AVERAGES AND VARIANCES FOR ANNUAL LEAVE

	July	August	September	October	November 1–14th
Averages (days)					
True values	3 814	4 393	3 539	1 714	0 439
Remembered	3.740	4 253	3 680	1.598	0 422
Discrepancies	−0 074	−0.140	+0 141	−0 116	−0 017
Average discrepancy as a percentage of the average true value	−2%	−3%	+4%	−7%	−4%
Variances					
True values σ_t^2	22 24	22 48	19 61	8 32	0 88
Remembered σ_r^2	24 47	25 04	21 83	8 26	0.92
Discrepancies σ_d^2	7 82	7 39	6 14	2.30	0.25
Ratio of the variance of the remembered values to the variance of the true values	1 10	1 11	1 11	0 99	1 05
Standard error of discrepancies	0 134	0 131	0 119	0 073	0.024
Correlation between discrepcies and true values ρ_{dt}	−0 212	−0 187	−0 179	−0 270	−0 216

Discrepancy = Remembered Value − True Value.

What would be the effect of altering the number of months? Our plans for collecting some evidence on these points fell through with the suspension of the Survey of Sickness and we shall have to wait for another opportunity to collect further data Before leaving the subject we will give in the next section some other examples of questions that have been used.

4 SOME MEMORY QUESTIONS

Experience with memory questions has made us realise the danger that errors arising from other causes may be wrongly attributed to memory It is important to remember that, to the informant, any recent medical consultation is interesting, not just the one occurring "yesterday" or in the "last seven days " The importance of this, and

other factors, is brought out in the following examples:

Example (i)

Did you consult a doctor yesterday, that is on ________? Yes ______ Y No ______ X
(WRITE IN THE DAY OF THE WEEK)
IF YES (Y), for what? ________________

This question was used experimentally at various times during 1951 as the final question in the Survey of Sickness interviews. It started life without the concluding words, "that is on__________," and also without the dependent question "for what?" These were added after checks had been made by mail and by reinterviewing. Checking the answers to memory questions by re-interview or by mail cannot of course be exact, due to the interval of time that must elapse before the check is made, but these checks did reveal some errors We found that the interviewer was sometimes dating her questionnaires wrongly. We also found that a few informants were saying "yes" when a consultation had occurred on the day of the interview The addition of "that is on __________" was an attempt to deal with this There were also a number of cases where the code Y (Yes) had been rung where the informant later insisted there had been no consultation on that day, or any day near to it (they confirmed however that interviews had been made on the correct days) However we found no code X (No) which should have been code Y This could occur if they were recording errors, and if it was equally likely for a code Y to be wrongly rung for a code X, as an X for a Y. For suppose such a recording error occurred once per five hundred questionnaires, then with a sample of 4,000 we might expect to find no X wrongly rung but eight Y's wrongly rung (the number of persons who would have consulted a doctor the day before the interview would be between 40 and 80) This would appear to be the reason for these errors "*IF YES (Y)*, for what?", even if the answers are never themselves used, would appear to be the answer to this problem. Indeed this use of dependent open questions, where the actual answers may not even be used, seems to be of general applicability in reducing the recording errors on precoded questions

Example (ii)

Have you consulted a doctor during the last seven days, that is from last
____
(WRITE IN AND NAME DAY)

to yesterday, inclusive? Yes ______ Y No ______ X
IF YES (Y) which days? ________________
(NAME DAYS AND CHECK THAT IN RIGHT PERIOD)

This question was also used experimentally during 1951 in the Survey of Sickness interview It also started life without the concluding words, "that is____________ " Checks revealed that the last seven days were not sufficiently well-defined There were cases of informants including the day of interview as well as the seven days intended, and cases where this day was counted as one of the seven days To cope with this the concluding phrase was added, and the interviewers were asked to check that the days named in answer to the dependent question were in the right period Our checks suggested that with this memory period there was then little or no memory error

Example (vii)

FIRST, I WOULD LIKE TO ASK YOU ABOUT THE LAST SEVEN DAYS, THAT IS FROM LAST ________ ____________ TO YESTERDAY, INCLUSIVE

WRITE IN AND NAME DAY
IT WILL BE THE SAME DAY OF THE WEEK AS THE DAY OF INTERVIEW

	February		March		
Sunday		24	—	2	9
Monday		25	—	3	10
Tuesday		26	—	4	11
Wednesday		27	—	5	12
Thursday		28	—	6	13
Friday		29	—	7	14
Saturday	23	—	1	8	15

1. Have you consulted a doctor for any reason at all during the last seven days, that is from last ____________ to yesterday, inclusive? (INCLUDE ANY CONSULTATIONS ALREADY MENTIONED.)

Yes____Y
No____X

IF YES (*Y*) How many times? ____
ASK FOR EACH CONSULTATION IN TURN

IF NO (*X*) When did you last consult a doctor for any reason at all?
____
GO ON TO QUESTION 2

	VISIT A	VISIT B	VISIT C
(a) Which days?	______	______	______
(b) Did you see your usual doctor on this visit?	Yes____Y	Yes____Y	Yes____Y
IF NO Who did you see?	Who seen____	Who seen____	Who seen____
	____□	____□	____□
(c) Was this under the National Health Scheme?	National Health___X	National Health___X	National Health___X
	Private____O	Private____O	Private____O
Other questions followed			

These questions were used in March 1952 in an inquiry [4] carried out for the Committee on General Practice of the Central Health Services Council They were added to the end of the Survey of Sickness questionnaire. It will be seen that the memory period chosen was the last seven days prior to interview. This period was chosen partly as a result of our earlier experiments but mainly because, on this inquiry, as distinct from the Survey of Sickness itself, far greater detail was required about what occurred during a sample of consultations. Other dependent questions followed the three printed above, asking for such

details as the prescriptions issued and the certificates given. It was mainly to prevent confusion between what had occurred at different consultations that these questions were confined to consultations made in the last seven days.

It will be seen that the informant was told the memory period twice, once in the introduction and once in the main question itself. The interviewer had to write in the day of the week in each case and, as is our general practice with memory questions, was able to refer to a small calendar printed alongside the question. Both Yes and No answers were followed by dependent questions, the Yeses being asked how many times and on which days. Where only one consultation was made, the day of the week was entered in column A and the other dependent questions followed. Where more than one consultation had occurred the days of the week were entered at the heads of the columns and the interviewer asked all the questions, first about, say, "the consultation on Tuesday" and then repeated the questions for "the consultation on Thursday", and so on. The arrangement of the questions made it easy for the interviewer to check that the same medicine or certificate was not wrongly attributed to two consultations.

No attempt was made on this occasion to deal with the problem, already mentioned in the first section of this paper, which arises because people are not necessarily found at home when the interviewer calls the first time. However, these questions were to have been asked with the Survey of Sickness in the two following months of April and May and it was hoped to study this effect in the last month. Our plans involved splitting the sample into two, confining the interviewing for each part to one week and recording the date of each call at every address. This meant in practice that the last seven days would always include the day of the first call. In this way we hoped to learn something of the effect of any relationship between seeing the doctor and being at home when the interviewer called. Unfortunately, the suspension of the Survey of Sickness also resulted in the cancellation of the remaining portions of this inquiry.

5. CONCLUSION

The examples given in this article are confined to one type of survey dealing with sickness. Different results may well be obtained with different subjects Furthermore, only single interview surveys have been considered. Longer memory periods would be possible, no doubt, if an interview were made at the beginning, as well as the end of the memory period, or if some form of record-keeping was adopted. Nevertheless it

does illustrate the factors to be kept in mind in dealing with memory questions.

In considering the length of memory period to be chosen it is necessary to remember that there is likely, with many variables, to be a positive correlation over consecutive periods of time, with the result that the gain obtained by increasing the length of memory period may be less than might be expected. The loss due to memory takes two forms. The variance of the remembered values is likely to exceed the variance of the true values and there is likely to be a residual bias The predominant type of error in the experiment reported was one of transference, where the informant transfers the events both in and out at both ends of the selected period This suggest that a period ending on the day before the interview is to be preferred, wherever possible, since this closes the door to one set of transfers An interview made at the beginning of the period would no doubt help to prevent transfers at that end of the period. Perhaps a letter might serve the same purpose.

The choice of such a memory period means that it can be fixed in length, but not in time, since not all the interviews, or even a subsample of them, can be made on a given day Apart from this obvious disadvantage that the periods do not all commence and end on the same day, there is another disadvantage which is perhaps not always appreciated It arises because not all the informants are found at home at the first call, and it is possible for some relationship to exist between the variable being studied and the likelihood of the informant being found at home on a given day Little information is yet available on this point.

Our experience with a number of questions has made us realize the importance of carefully defining, and redefining, the memory period for the informant It has to be remembered that it is often the event, and not when it happened, which appears important to the informant

The suspension of the Survey of Sickness brought to an end this particular line of inquiry and it seems an appropriate time to describe what has been learned I am well aware that this article raises more questions than it answers but it has been written in the hope that it will stimulate others to publish their data on this important subject. Little has been published. Of the very sparse literature on the subject, perhaps the most interesting is an article by Mauldin and Marks [5], though they, too, give little quantitative data

Finally, I should like to acknowledge the help that I have received from my colleagues both present and past.

REFERENCES

[1] Monthly Bulletin of the Ministry of Health, June 1944, 93
[2] Monthly Bulletin of the Ministry of Health, September 1945, 198
[3] *Sickness in the Population of England and Wales in 1944–47*, H.M S.O., London, 1949.
[4] Gray, P G. and Cartwright, A , "General practice under the national health service," Social Survey Report, June 1952.
[5] Mauldin, W. P., and Marks, E S., "Problems of response in enumerative surveys," *American Sociological Review*, 15 (1950), 649–57

MEASURING THE ERROR OF EDITING THE QUESTIONNAIRES IN A CENSUS

SVEIN NORDBOTTEN
Central Bureau of Statistics, Oslo

This paper describes an attempt to measure the relative errors in totals and their influence on size distributions due to the questionnaire editing in a census. The experiment described was performed in connection with the 1953 Industrial Census in Norway The results indicate that the errors are relatively small and their influence on distributions insignificant The gains of a more thorough editing would probably be small and efforts might perhaps be better spent by improving the census procedure by other means.

INTRODUCTION

THE results of a census are often published without any statement about the quality of the statistics Both the professional statistician and the consumers of statistics would obviously profit by a greater knowledge of the accuracy of the statistics

In a large census there may arise a great number of errors which may have an important influence on the results Deming [2] gives an extensive list of possible errors The professional statistician will be interested not only in a quantitative measure of the different error components, but also in their causes in order to be able to make more efficient census designs in the future

The error components may be classified in two main groups:

(a) Random errors which arise because some element of random selection has been introduced in the census procedure

(b) Non-random errors which are due to factors such as bad design, collection, editing, processing, etc. This group may be subdivided into subclasses

Different approaches have been tried in evaluating total or relative non-random errors and their components in surveys and censuses. One which seems to be successful [3], is to check a small probability sample and compare census and sample results item-by-item

In connection with the 1953 Industrial Census in Norway an experiment was done along these lines in order to try methods for evaluating quantitative measures for the editing errors and testing their influence on distributions of establishments.

DESCRIPTION OF THE EXPERIMENT

The purpose of the experiment presented in this paper was to investigate the relative editing error of statistics describing a mass of small manufacturing establishments, and the extent to which these errors influence certain distributions of the establishments.

The experiment was limited to Industrial Census questionnaires from six to seven thousand establishments, edited by six persons. Each day a five per cent systematic sample with random starts was drawn from the questionnaires edited that day by each person. The sample was then submitted to a second and very thorough control editing, independent of the original edit, with respect to employment, wage expenditures and value of production The control editing was done by one specially instructed person

The six persons performing the first editing were instructed not to contact the respondents for any supplementary information. When no answer was given or when an answer was supposed to be wrong, they were to make a rough estimate of the characteristic. In this way, 6620 questionnaires were edited. On the other hand, the one person performing the control editing was instructed to use all means to get correct information about employment, wage expenditures, and the value of production For 103 of the 331 sample questionnaires, one or more requests for further information were sent

These definitions were used·

(*a*) The individual error of a characteristic is the difference of the result obtained by the first and the second editing of a questionnaire.

(*b*) The relative error of a characteristic is the sum of all individual errors divided by the sum of the result obtained if all questionnaires were submitted to the control editing

The following notation was used:

n = total number of establishments in the sample.

N = total number of establishments in the population.

f = sampling fraction.

L = number of persons performing the editing.

n_h = number of establishments in the sample from the hth person.

x_{hi} = individual error on questionnaire i in the sample from the hth person.

y_{hi} = value of the characteristic for questionnaire i in the sample from the hth person after control editing.

r_h = estimate of the relative error of the population edited by the hth person

r = estimate of the relative editing error.

s^2 = estimate of the variance of r

$a_i(k) = 1$ if the ith establishment in the sample is classified in class k in the first editing,
$= 0$ otherwise.

$b_i(k) = 1$ if the ith establishment in the sample is classified in class k in the control editing,
$= 0$ otherwise.

$c_i(k) = 1$ if the ith establishment in the sample is classified in class k in both editings,
$= 0$ otherwise

M = number of classes

The first aim was to estimate the relative errors of the characteristics These errors were estimated by means of the ratio estimator

$$r = \sum_h^L \sum_i^{n_h} x_{hi} \Big/ \sum_h^L \sum_i^{n_h} y_{hi}. \tag{1}$$

This estimate is biased On the other hand, it is consistent and the bias in this case is expected to be insignificant. When the number of questionnaires included in the population is as great as here and the sampling fraction is only five per cent, the estimate should have an approximate normal distribution if the sample is regarded as a stratified random sample. The variance of r was estimated by [1]

$$s^2 = (1 - f) \sum_h^L \left(\sum_i^{n_h} x^2_{hi} + r^2_h \sum_i^{n_h} [y^2_{hi} - 2r_h \sum_i^{n_h} x_{hi} y_{hi} \right) \Big/ \left(\sum_h^L \sum_i^{n_h} y_{hi} \right)^2. \tag{2}$$

To test the hypothesis that the editing errors had no influence on the distributions of the establishments by employment, wage expenditures, and value of production, the following variable was formed

$$K^2 = \sum_k^M (h_1(k) - h_2(k))^2 / \text{var}\,(h_1(k) - h_2(k)), \tag{3}$$

where

$$h_1(k) = \frac{N}{n} \sum_i^n a_i(k), \tag{4}$$

(5) $$h(k) = \frac{N}{n} \sum_{i}^{n} b_i(k), \qquad k = 1 \cdots M$$

and

(6) $$\operatorname{var}\,[h_1(k) - h_2(k)] = \frac{N}{n}\left(\frac{N-n}{N-1}\right) L\left[h_1(k) + h_2(k) - 2\frac{N}{n}\sum_{i}^{n} c_i(k)\right].$$

Assuming that the differences $h_1(k) - h_2(k)$, $k = 1 \cdots M$, are normally distributed, the variable K^2 will be approximately chi-square distributed with $M-1$ degrees of freedom [4].

We chose a one per cent level of significance and defined a value χ_0^2 such that the probability

(7) $$P(\chi_0^2 \geqq K^2;\, M-1) = 0\ 99$$

The hypothesis that the editing error does not influence the distribution of the establishments should be rejected when $K^2 > \chi_0^2$.

THE RESULTS OF THE EXPERIMENT

The estimates of the relative editing errors and their standard deviations computed from formulas (1) and (2) are given in Table 1.

TABLE 1

ESTIMATES OF THE RELATIVE EDITING ERRORS AND THEIR STANDARD DEVIATIONS

Characteristic	Relative editing error	Standard deviation
Employment	0 0273	0 0126
Wage expenditures	0 0248	0 0128
Value of production	0 0172	0 0114

The table indicates that the relative editing errors are rather small in spite of the imperfect editing All of them seem to be positive. Even though there is no significant difference, it is also interesting to note that the relative error is lowest for the value of production. With a confidence coefficient of 0.99, the maximum relative error of employment is 6 5%, of wage expenditures 6.3%, and of the value of production 5 1%

The relative distributions of the sampled establishments by employment, wage expenditures and value of production are given in Table 2.

TABLE 2

RELATIVE DISTRIBUTIONS OF SAMPLED ESTABLISHMENTS

Class	Distribution after first editing: Employment	Wage expenditures	Value of prod.	Distribution after second editing: Employment	Wage expenditures	Value of prod.
1	8.5%	46 8%	27 1%	10 0%	48.1%	29 0%
2	8.2	16 6	16.3	10 6	16.6	16.9
3	34.4	12.1	12 6	31 1	11 5	11.1
4	14.2	4.5	9 1	14 8	4.5	9.1
5	10 6	4.5	7 6	10 8	4 5	7 6
6	7 8	3.4	6.9	7 3	2 8	6.0
7	7 6	5 1	3.0	7.3	5 1	3.0
8	3 0	7 3	3.0	3.0	6 9	3 0
9	5 7	—	3.6	5.1	—	3 6
10	—	—	4 8	—	—	4.7
11	—	—	6.0	—	—	6.0

The hypothesis that the distributions of the establishments by employment, wage expenditures, and value of production after the first editing are equal to the distributions by the same characteristics which would have been obtained if the control editing had been complete, was tested with the following results

TABLE 3

Distribution by	K^2	χ_0^2	Significant deviation
Employment	19.529	20 090	No
Wage expenditures	4 363	18 475	No
Value of production	13 143	23 209	No

CONCLUSION

The results of the experiment described in this paper indicate that the effects of errors in editing questionnaires on statistics for a population of *small* establishments were small. The tests did not reject our hypothesis that there are no significant deviations between the distributions obtained by the first editing procedure and those which would have been obtained if all questionnaires were submitted to a control editing. The accuracy gained by a thorough and expensive editing of

the questionnaires is small and may perhaps be obtained cheaper by improving the census procedure by other means.

REFERENCES

[1] Cochran, William G., *Sampling Techniques*. New York: John Wiley and Sons, 1953, 118–19.

[2] Cramér, Harald, *Mathematical Methods of Statistics*. Princeton: Princeton Mathematical Series, 1946, 378–87.

[3] Deming, William Edwards, *Some Theory of Sampling*. New York: John Wiley and Sons, 1950, 24–30.

[4] Eckler, A Ross, "Extent and character of errors in the 1950 Census," *The American Statistician*, 7 (1953), No. 5, 15–19

THE RELATIONSHIP OF HOUSING PRICES AND BUILDING COSTS IN LOS ANGELES, 1900–1953*

ROBERT M WILLIAMS
University of California (Los Angeles)

Series measuring residential construction costs and asking prices for single family dwellings in Los Angeles both had steeply rising trends of approximately equal slope from 1900 to 1953. Both series described clearly defined cyclical fluctuations of wide amplitude However, asking prices, which are indicative of sales values, fluctuated more widely than construction costs The ratio of asking prices to construction costs traced a cyclical pattern of large amplitude with troughs occurring in 1900, 1919, 1934, 1942, and 1950. Except for the immediate postwar period, 1945–1947, this ratio fluctuated less widely since 1934 than in the earlier part of the period analyzed

IN A recent article in this *Journal* Blank compared secular movements in average house prices in 22 cities with a construction cost index.[1] His purpose was to test the validity of the common practice of substituting construction cost indexes—which are generally available—for market price series—which are generally nonexistent—for property assessment and other purposes He concluded that, for the series examined, "the construction cost index measures with quite reasonable accuracy the secular movement of house prices."[2] He showed, however, that short-run divergences between the two series at one point exceeded 30 per cent, although the difference was less than 10 per cent during most of the period from 1890 to 1934 Hence, for short-term analysis, "some margins of error are involved in using the cost index as an approximation of a price index "[3]

The present article will examine fluctuations in housing prices and building costs in Los Angeles from 1900 to 1953. This analysis supports Blank's conclusions concerning both the similarity in underlying trends in housing prices and building costs and the significant divergence between the two series from year to year It will be shown, however, that the Los Angeles housing price index had much greater amplitude and traced a considerably more regular cyclical pattern than Blank's

* Much of the material contained in this article is derived from a study of *Residential Construction in Los Angeles, 1900–53*, under the direction of the Bureau of Business and Economic Research, University of California, Los Angeles.

[1] David M Blank, "Relationship Between an Index of House Prices and Building Costs," Vol. 49 (1954), 67–78

[2] *Ibid.*, p 78

[3] *Ibid.*, p. 78

22-city price series. Finally, reasons will be suggested for the differences in cyclical amplitude and regularity between the Los Angeles housing price series and that for the 22 cities combined Before examining the relationship between housing prices and construction costs in more detail, the housing price series developed for Los Angeles will be described.

AN INDEX OF ASKING PRICES FOR SINGLE FAMILY DWELLINGS IN LOS ANGELES

Mr Blank's 22-city index of housing prices was based on the owner's estimate of the value of his property in 1934 and his recollection of the price paid when the property was acquired[4] Correcting the data for both normal depreciation and appreciation from improvements, he constructed an index of housing values from 1890 to 1934.

The Los Angeles series is based on asking prices for single family dwellings. Asking-price data were selected quarterly from classified advertisements for 5- and 6-room houses built on standard sized lots. It is recognized that asking prices differ from actual selling prices by an average margin which may vary cyclically.[5] Moreover, houses advertised for sale may not be representative of the entire population of single family dwellings, a population which itself is constantly changing in composition.

This asking-price index was compared with a market-price series available from 1940 to 1953. The patterns of behavior in the two series were remarkably similar during this 13-year period in which average prices increased approximately 200 per cent and building costs nearly as much[6] The two independently constructed price series not only had similar secular trends, but also had nearly identical year-to-year fluctuations This parallel movement since 1940, plus the fact that the asking-price series displays the expected close correlation with residential construction for the period from 1900 to 1940, suggests that a carefully constructed asking-price index can be substituted for market prices.

The close relationship between the asking-price index and new dwelling units authorized in Los Angeles County is shown in Table 1. Both series increased rapidly after 1900, turned down briefly under the im-

[4] Data for Blank's analysis were derived from the *Financial Survey of Urban Housing* (Washington, D C U S Department of Commerce, 1937)

[5] A preliminary analysis of data supplied by multiple-listing services in the Los Angeles and San Francisco areas indicates that average sales price has varied in recent years between 90 and 96 per cent of the list or asking price As might be expected, there is some evidence that the spread between sales and asking price is greater when values are declining than when they are increasing

[6] See Williams, Robert M., "An index of asking prices for single family dwellings," *The Appraisal Journal*, 22 (1954), 33–8, for a description of the asking-price and market-price indexes for Los Angeles.

TABLE 1

ASKING-PRICE INDEX AND DWELLING UNITS AUTHORIZED IN LOS ANGELES COUNTY, 1900–1953

Year	Asking-Price Index (1940=100)	Dwelling Units Authorized (in thousands)	Year	Asking-Price Index (1940=100)	Dwelling Units Authorized (in thousands)
1900	36	4.1	1930	122	21.5
1901	43	5.2	1931	111	12.0
1902	54	7.6	1932	89	5.4
1903	61	10.7	1933	76	4.8
1904	63	11.8	1934	75	3.9
1905	66	13.1	1935	86	8.9
1906	70	14.5	1936	92	19.3
1907	74	9.2	1937	107	24.2
1908	70	6.3	1938	110	30.5
1909	70	10.3	1939	104	39.2
1910	69	18.2	1940	100	44.5
1911	72	18.9	1941	107	47.7
1912	72	26.1	1942	113	26.6
1913	73	18.4	1943	130	14.6
1914	69	12.2	1944	156	32.6
1915	63	8.5	1945	188	27.7
1916	60	6.1	1946	278	67.0
1917	60	4.3	1947	281	67.2
1918	65	3.4	1948	277	79.1
1919	76	9.2	1949	240	69.4
1920	116	20.3	1950	249	96.6
1921	134	35.4	1951	282	62.1
1922	143	54.3	1952	296	78.8
1923	180	82.9	1953	300	84.9
1924	190	53.4			
1925	182	41.5			
1926	163	43.3			
1927	152	38.8			
1928	142	31.9			
1929	130	27.3			

pact of the 1907 recession in general business activity, again increased to 1912 or 1913, and then declined to a lower turning point in 1917 or 1918. Both series increased rapidly in the early twenties, reached an upper turning point in 1923 or 1924, and then declined until 1934. After

1934, both series again increased. The asking-price index, however, declined in 1939 and 1940, while dwelling units authorized increased each year until 1942, when wartime building restrictions were imposed.

Space does not permit more detailed analysis of the relationship between these two series This brief description of parallel movement is presented to suggest that the asking-price index behaved as one would expect a market-price series to behave, because the rate of new residential construction would be expected to have a close relationship to the market prices of existing dwellings Hence, because asking prices and rate of residential building are closely related, it appears that asking prices can be substituted for market prices in analyzing fluctuations in residential real estate series

THE RATIO OF ASKING PRICES TO BUILDING COSTS

The index of asking prices for single family residences in Los Angeles from 1900 to 1953 is presented in column 1 of Table 2 and an index of residential construction costs in column 2. The cost series was based on the Boeckh index of building costs for frame houses in Los Angeles from 1913 to 1953. Because no local cost series is available for earlier years, Blank's national index of construction costs was used from 1900 to 1913 This substitution is probably satisfactory, since the behavior of the national cost index in this period closely paralleled the average permit value per dwelling unit authorized in Los Angeles, a series which fluctuates with construction costs except in periods of rapid cost or income change It should be pointed out that actual building costs have greater cyclical fluctuation than building-cost indexes This is because the latter largely exclude contractors' profits, which vary much more than wage and material costs Hence, using a building cost index exaggerates somewhat the actual cyclical divergence of market values and replacement costs.

As shown in Chart 1, the secular trends of the Los Angeles price and cost series are quite similar—as Blank found to be the case for the national series The cost index increased nearly 500 per cent from 1900 to 1953. The price index increased somewhat more in this period, but the difference is not significant. Rather, it reflects the fact that market prices were depressed in 1900, while in 1953 residential values were at record levels

Although their trends are alike, the two Los Angeles series show considerable divergence in the short run. Whereas Blank's price series fluctuated above and below the cost series in a more or less erratic fashion, the Los Angeles price index fluctuates widely about the cost index in a clearly defined pattern.

TABLE 2

ASKING-PRICE AND CONSTRUCTION-COST INDEXES IN LOS ANGELES, 1900–1953 (1940 = 100)

Year	Asking-Price Index	Construction-Cost Index	Ratio of Price to Cost Index	Year	Asking-Price Index	Construction-Cost Index	Ratio of Price to Cost Index
	(1)	(2)	(3)		(1)	(2)	(3)
1900	36	44	82	1930	122	90	136
1901	43	43	100	1931	111	84	132
1902	54	45	120	1932	89	76	117
1903	61	46	133	1933	76	79	96
1904	63	46	137	1934	75	83	90
1905	66	48	138	1935	86	83	104
1906	70	53	132	1936	92	88	105
1907	74	55	134	1937	107	97	110
1908	70	53	132	1938	110	102	108
1909	70	55	127	1939	104	97	107
1910	69	57	121	1940	100	100	100
1911	72	56	129	1941	107	112	96
1912	72	58	124	1942	113	124	91
1913	73	56	130	1943	130	129	101
1914	69	54	128	1944	156	138	113
1915	63	56	112	1945	188	145	130
1916	60	59	102	1946	278	161	173
1917	60	66	91	1947	281	199	141
1918	65	79	82	1948	277	225	123
1919	76	103	74	1949	240	211	114
1920	116	132	88	1950	249	231	108
1921	134	108	124	1951	282	247	114
1922	143	102	140	1952	296	251	118
1923	180	111	162	1953	300	256	117
1924	190	107	178				
1925	182	99	184				
1926	163	98	166				
1927	152	96	158				
1928	142	95	149				
1929	130	97	134				

Source *Column 1*—Index derived from quarterly average of median asking prices for 5- and 6-room houses in classified advertisements.

Column 2—1900–1912 based on Blank's national index of residential construction costs See David M Blank, "Relationship between an Index of House Prices and Building Costs," *Journal of the American Statistical Association*, Vol 49 (1954), p 76, Table 4 for sources

1913–1953 Boeckh building-cost index for frame residences as reported in *Housing Statistics Handbook* and *Engineering News-Record*

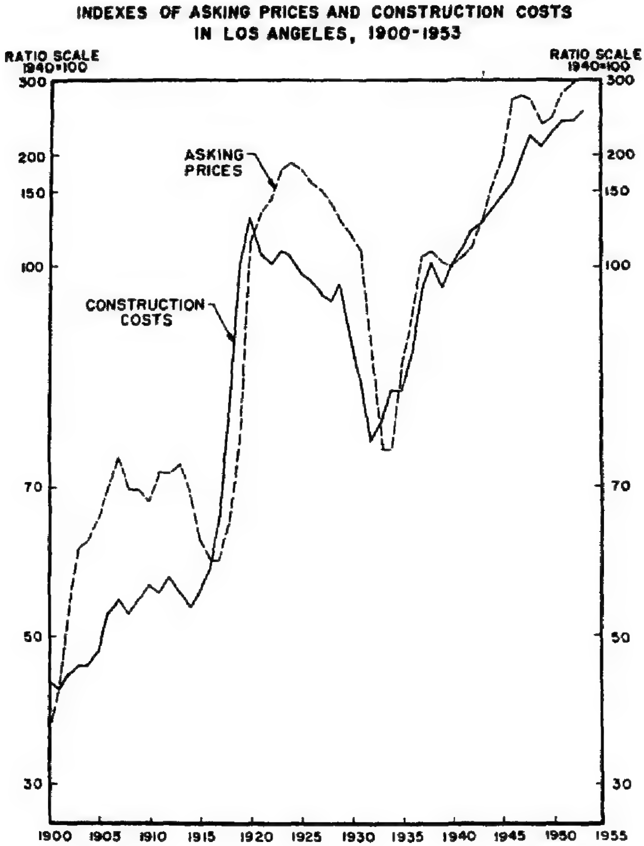


CHART I

Between 1900 and 1934, two complete cycles of wide amplitude are discernible in the ratio of prices to costs (see Chart 2). These cycles, which correspond almost exactly to those in new residential construction, had periods of 19 and 15 years, respectively, from trough to trough Their amplitudes, measured as the difference between the minimum and maximum values of the ratio in each cycle, were 64 and 110 percentage points, respectively. After 1934, the ratio fluctuated less widely, except for the immediate postwar years, increasing from 1934 to 1937 and then declining until 1942 New residential construction, however, continued to increase each year from 1934 to 1941 (see Table 1)

RATIO OF ASKING PRICES TO CONSTRUCTION COSTS
IN LOS ANGELES, 1900-1953

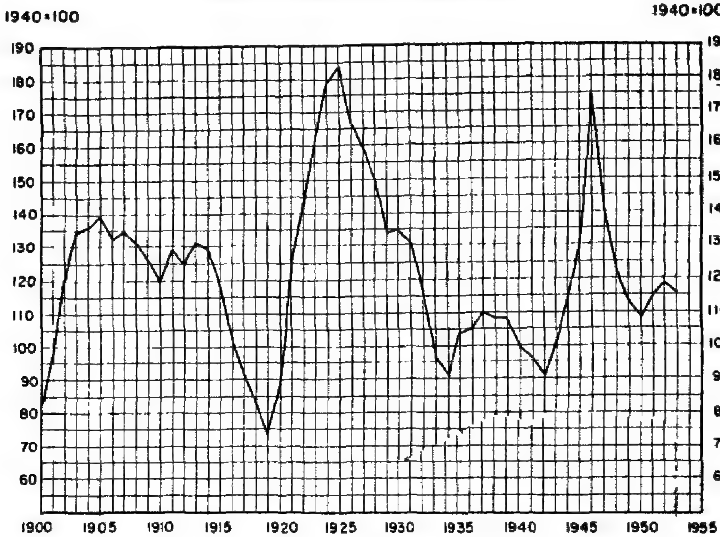


CHART II

Several factors may have contributed to differences in behavior between Blank's 22-city price series and that for Los Angeles. One is Blank's method of correcting his price series for depreciation and capital improvement by making a uniform adjustment over the cycle This may have produced a more stable price series than would otherwise have resulted Moreover, in combining data for 22 cities differences in cyclical turning points would blur the cyclical pattern existing in series for individual cities. Finally, Los Angeles has grown very rapidly—more than doubling in population in each of the two decades beginning in 1900 and 1920. These upsurges in population coincided with sharp increases in rents, housing values, and rate of new construction 'he available data for other cities, however, including B L.S. series on 'ents and new dwelling units authorized, indicate that wide fluctuations n real estate series have occurred in more slowly growing areas also

CONCLUSIONS

Wide fluctuations occurred in the ratio of housing prices to construc-:ion costs in Los Angeles from 1900 to 1934 and also after World War I. *Hence, a cost index could not have been used satisfactorily to* neasure changes in housing market values in these periods.

A TEST OF A LINEAR FUNCTION OF THE DEVIATIONS BETWEEN OBSERVED AND EXPECTED NUMBERS*

WILLIAM G COCHRAN
The Johns Hopkins University

I. INTRODUCTION

As is well known, the χ^2 test of goodness of fit is not directed against any specific pattern of the deviations $(f_i - m_i)$ of the observed frequencies f_i from the expected frequencies m_i. For this reason, the χ^2 test is sometimes insensitive in detecting a failure of the null hypothesis There are, however, a number of alternative or supplementary tests that may be used when it is possible, from the nature of the problem, to predict the type of alternative hypothesis that is most likely to hold if the null hypothesis fails. These tests include a comparison of the variances, or of the third and fourth moments, of the observed and theoretical distributions, and various ways of breaking down χ^2 into components [1].

One additional test of this kind is obtained by selecting any linear function of the deviations,

$$L = \sum g_i(f_i - m_i),$$

where the g_i are numbers, chosen in advance by the person making the test, in such a way that L will be sensitive to the alternative hypothesis that is thought most likely to hold. By suitable assignment of the numbers g_i, the criterion L can be made responsive to any anticipated pattern of deviations, either in their signs or in their magnitudes In particular, if all but one g_i are put equal to zero we obtain the test of an individual deviation This paper describes how to make an approximate test of significance of the value of L. The test is approximate in roughly the same sense in which the goodness of fit χ^2 test is itself approximate, i.e., the test is strictly valid as an asymptotic result when the expectations become large The theory of the test will be presented first, followed by several illustrative examples

II EXPECTATIONS GIVEN IN ADVANCE

For simplicity, we consider first the case in which the expectations are completely specified by the null hypothesis, so that there are no

* Work assisted by a contract with the Office of Naval Research, Navy Department. Department of Biostatistics paper No. 301

unknown parameters to be estimated This is not the common situation, but it does occur, for instance, in testing Mendelian inheritance where the binomial p is given by theory

Let the known expectations be denoted by the symbols M_i and let

$$N = \sum M_i = \sum f_i$$

be the total size of sample On the null hypothesis, the observed frequencies f_i follow a multinomial distribution with the following properties:

Mean:

$$E(f_i) = M_i \tag{1}$$

Variance:

$$V(f_i) = M_i \left\{1 - \frac{M_i}{N}\right\} \tag{2}$$

Covariances:

$$\text{Cov}\,(f_i, f_j) = -\frac{M_i M_j}{N} \quad (i \neq j) \tag{3}$$

Then, taking

$$L' = \sum g_i(f_i - M_i),$$

we have, on the null hypothesis,

$$E(L') = 0$$

$$\begin{aligned} V(L') &= \sum_i g_i^2 V(f_i) + 2 \sum_{i<j} g_i g_j \,\text{Cov}\,(f_i, f_j) \\ &= \sum_i g_i^2 M_i \left\{1 - \frac{M_i}{N}\right\} - \frac{2}{N} \sum_{i<j} g_i g_j M_i M_j \\ &= \sum g_i^2 M_i - \frac{1}{N} \left(\sum g_i M_i\right)^2 \end{aligned} \tag{4}$$

This is an old result, and is exact for any size of sample [3, §55]

Further, as N becomes large, with fixed p's, so that the M_i become large, the multinomial distribution of the f_i tends to a multivariate normal distribution [2, §30 1], and L' tends to be normally distributed Hence, the test of significance is made either by treating $L'/\sigma(L')$ as a normal deviate, or by treating

$$\chi_{L'}^2 = \frac{L'^2}{V(L')}$$

as χ^2 with 1 degree of freedom

III. ONE-PARAMETER ESTIMATION

When the expectations m_i are estimated from the sample, the situation is more complex, and the formula to be given for $V(L)$ is valid only when the expectations are large. Suppose that the M_i are known functions of a single unknown parameter θ, of which the sample estimate is $\hat{\theta}$. Maximum likelihood estimation, or some asymptotically identical method, is assumed The symbol L will denote the linear function when the expectations are estimated, while L' will be used, as in Section II, when the expectations are known

We now have

$$L = \sum g_i(f_i - m_i)$$

where the m_i are the values taken by the M_i when $\theta = \hat{\theta}$ We will first find the variances and covariances of the $(f_i - m_i)$, noting that the m_i are now functions of the sample observations.

To save space, two standard results in the theory of maximum likelihood estimation will be quoted These results, and most results that follow from them, are valid apart from terms that can be neglected when the expectations are large The symbol $\doteq$ denotes an equation of this type.

The first assumed result is

$$\hat{\theta} - \theta \doteq \frac{1}{I} \sum \frac{(f_i - M_i)}{M_i} \frac{\partial M_i}{\partial \theta}, \tag{5}$$

where

$$I = \sum \frac{1}{M_i} \left(\frac{\partial M_i}{\partial \theta} \right)^2 \tag{6}$$

is Fisher's "amount of information," and is the inverse of the asymptotic variance of $\hat{\theta}$. (In expositions of maximum likelihood theory, the result (5) usually appears in a slightly different form. For instance, in Cramér [2], equation (33 3.4), the right side of (5) has an additional denominator which, as Cramér shows, converges in probability to 1 when the expectations become large.)

The second assumed result is

$$m_i - M_i \doteq (\hat{\theta} - \theta)\frac{\partial M_i}{\partial \theta}, \tag{7}$$

this being obtained by the first term of a Taylor expansion. Equations (5) and (7) require certain restrictions on the forms of the functions M_i and their derivatives (cf Cramér, loc. cit.), but I believe that these are satisfied in any of the common applications.

Substitution of the value of $(\hat{\theta}-\theta)$ from (5) into (7) gives

$$m_i - M_i \doteq \frac{1}{I}\frac{\partial M_i}{\partial \theta}\sum_j \frac{(f_j - M_j)}{M_j}\frac{\partial M_j}{\partial \theta}.$$

Hence,

$$\begin{aligned} f_i - m_i &= (f_i - M_i) - (m_i - M_i) \\ &\doteq (f_i - M_i) - \frac{1}{I}\frac{\partial M_i}{\partial \theta}\sum_j \frac{(f_j - M_j)}{M_j}\frac{\partial M_j}{\partial \theta} \end{aligned} \tag{8}$$

This is the key equation It expresses (f_i-m_i) as a linear function of the deviations (f_j-M_j) from the *true* expectations. Since the variances and covariances of these latter deviations are known from equations (2) and (3), we can now find the variance of (f_i-m_i), or of any linear function L of these deviations Equation (8) also implies that, to the present order of approximation, $E(L)=0$ when the null hypothesis holds

Instead of proceeding directly, we shall follow a different route that appears to simplify the algebra. The right side of equation (8) may be interpreted as the deviation of (f_i-M_i) from its linear regression on the variate

$$X = \sum \frac{(f_j - M_j)}{M_j}\frac{\partial M_j}{\partial \theta}. \tag{9}$$

To see this, we have from (2) and (3)

$$\operatorname{Cov}\{(f_i - M_i), X\} = \frac{\partial M_i}{\partial \theta} - \frac{M_i}{N}\sum_j \frac{\partial M_j}{\partial \theta} = \frac{\partial M_i}{\partial \theta}, \tag{10}$$

since $\sum M_j = N$, so that its derivative vanishes Similarly, we find

$$V(X) = \sum \frac{1}{M_i}\left(\frac{\partial M_i}{\partial \theta}\right)^2 = I. \tag{11}$$

Thus the regression coefficient of (f_i-M_i) on X is

$$b_i = \frac{1}{I} \frac{\partial M_i}{\partial \theta}.$$

Hence, equation (8) may be rewritten as

$$(f_i - m_i) = (f_i - M_i) - b_i X.$$

If

$$L' = \sum g_i(f_i - M_i),$$

it follows that

$$L = \sum g_i(f_i - m_i) = L' - bX,$$

where $b = \sum g_i b_i$ is the regression coefficient of L' on X.

Hence the variance of L is equal to the variance of the deviations of L' from its linear regression on X, i.e.,

$$V(L) = V(L') - \frac{[\mathrm{Cov}\,(L', X)]^2}{V(X)}. \tag{12}$$

But, from (10) and (11),

$$\mathrm{Cov}\,(L', X) = \sum g_i \frac{\partial M_i}{\partial \theta} : V(X) = I$$

This gives, finally,

$$V(L) \doteq (\sum g_i^2 M_i) - \frac{(\sum g_i M_i)^2}{N} - \frac{1}{I}\left(\sum g_i \frac{\partial M_i}{\partial \theta}\right)^2 \tag{13}$$

In practice, we substitute the computed expectations m_i in place of M_i. For testing a single deviation, we have

$$\hat{V}(f_i - m_i) \doteq m_i - \frac{m_i^2}{N} - \frac{1}{I}\left(\frac{\partial m_i}{\partial \theta}\right)^2, \tag{14}$$

where $\hat{V}$ denotes an estimated variance.

As in the case where the expectations are given, the test is made either by treating $L/s(L)$ as a normal deviate, or by treating $L^2/\hat{V}(L)$ as χ^2 with 1 degree of freedom.

Like the goodness of fit test, this test requires some restriction on the smallness of the expectations. The restriction needed will depend on the form of the function L. It might be safe to allow some expectations as low as 1 or 2 if these expectations receive relatively small weights g_i in computing L. An example of the exact small-sample dis-

tribution is presented in the Appendix. So far as it goes, it suggests that the normal approximation may work about as well as does the tabular χ^2 approximation in the ordinary goodness of fit test Pending further investigation, it seems well to follow the common rule that the minimum expectation should not be less than 5, particularly when a single deviation is being tested With a single deviation, it is advisable to apply a correction for continuity, by taking the normal deviate as

$$\frac{|f_i - m_i| - \frac{1}{2}}{s(L)}.$$

There is a kind of intuitive interpretation to the result that the variance of L equals the variance of the deviations of L' from its regression on X In Section II, when no parameters were being estimated, the deviations $(f_i - M_i)$ were subject to the single restriction,

$$\sum (f_i - M_i) = 0.$$

It is this constraint that introduces the negative covariance in equation (3) between f_i and f_j When a parameter is being estimated, the maximum likelihood equation imposes a *further* constraint, i e ,

$$\sum \frac{f_i}{M_i} \frac{\partial M_i}{\partial \theta} = 0,$$

which may be written

$$X = \sum \frac{(f_i - M_i)}{M_i} \frac{\partial M_i}{\partial \theta} = 0$$

Thus the equation of estimation may be regarded as fixing the value of X. This additional restraint leaves the observed frequencies less free to deviate from the theoretical frequencies, and may be expected to diminish the variance of L as compared with that of L'. It is not surprising that the appropriate variance is now the variance of the deviations from the regression on the quantity X that is constrained by the equation of estimation

IV. APPLICATION TO THE BINOMIAL AND POISSON DISTRIBUTIONS

Perhaps the most common applications of goodness of fit tests are those to problems in which the null hypothesis specifies either a binomial or a Poisson distribution Formulas for $V(L)$ appropriate to these cases can be obtained by substituting the appropriate expressions for the M_i and their first derivatives in (13).

The binomial and Poisson distributions have the common property that the equation of estimation makes the sample mean equal to the theoretical mean. By applying the "regression" argument, we can obtain a more general formula for $V(L)$, applicable to any discrete distribution in which the maximum likelihood estimate is the sample mean (or a function of it)

Let f_i denote the frequency of i "successes" ($i=0, 1, 2 \cdots$) We may take X as

$$X = \sum i(f_i - M_i),$$

since the equation of estimation makes this quantity zero

Then, it is easy to verify that

$$\begin{aligned}\operatorname{cov}(L', X) &= \sum i g_i M_i - \frac{(\sum g_i M_i)(\sum i M_i)}{N} \\ &= \sum g_i M_i (i - \mu)\end{aligned} \tag{15}$$

where μ is the mean of the theoretical distribution. Also

$$V(X) = N\sigma^2$$

where σ^2 is the variance of the theoretical distribution.

Hence, for the estimated variance of deviations from the regression,

$$\hat{V}(L) \doteq \sum g_i^2 m_i - \frac{(\sum g_i m_i)^2}{N} - \frac{[\sum g_i m_i (i - \hat{\mu})]^2}{N\hat{\sigma}^2} \tag{16}$$

where we have substituted sample estimates for the unknown theoretical values involved For a sample of N from the binomial $(q+p)^n$, we substitute

$$\hat{\mu} = n\hat{p} : \hat{\sigma}^2 = n\hat{p}\hat{q}.$$

For a sample of N from the Poisson, we substitute

$$\mu = \hat{\sigma}^2 = \bar{i} = \text{sample mean}.$$

V EXAMPLES

Example 1. Fisher [3] has analyzed Geissler's data on the distribution of number of boys in 53,680 German families of size 8 (see Table 1). The null hypothesis is that the number of boys follows the binomial

$$(q + p)^8$$

where p is the proportion of boys. $\hat{p}$, as estimated from the data, is 0.51468.

TABLE 1

NO. OF BOYS IN FAMILIES OF 8

No of boys i	No. of families: Observed f_i	Expected m_i	Deviations f_i-m_i	g_i	ig_i
0	215	165 22	+ 49 78	+1	0
1	1,485	1,401 69	+ 83 31	−1	−1
2	5,331	5,202 65	+128 35	+1	+2
3	10,649	11,034 65	−385 65	−1	−3
4	14,959	14,627 60	+331 40	+1	+4
5	11,929	12,409 87	−480 87	−1	−5
6	6,678	6,580 24	+ 97.76	+1	+6
7	2,092	1,993 78	+ 98 22	−1	−7
8	342	264 30	+ 77 70	+1	+8
	53,680	53,680 00			

Fisher noted, as is apparent in Table 1, an excess of families with very unequal numbers of boys and girls He also noted an apparent bias in favor of even numbers of boys. This bias shows up in the central values (2–6 boys). At the extremes, the effect is obscured by the excess of unequally divided families

To test whether there is an excess of families with even numbers of boys, we may take

$$g_i = +1 \ (i \text{ even}) \quad g_i = -1 \ (i \text{ odd})$$

We find

$$L = \sum g_i(f_i - m_i) = +1369\ 98$$

To apply formula (16) for $V(L)$, we compute

$$\begin{aligned} \sum g_i^2 m_i &= 53{,}680 \\ \sum g_i m_i &= 0\ 02 \\ \sum i g_i m_i &= 0.09 \\ \hat{\mu} &= 4\ 1174 \\ \hat{\sigma}^2 &= 8\hat{p}\hat{q} = 1\ 9982 \\ N &= 53{,}680. \end{aligned}$$

Now apply formula (16)

$$\hat{V}(L) \doteq \sum g_i^2 m_i - \frac{(\sum g_i m_i)^2}{N} - \frac{[\sum g_i m_i (i - \hat{u})]^2}{N\hat{\sigma}^2}.$$

It is clear that the two subtraction terms are entirely negligible, so that

$$\hat{V}(L) \doteq \sum g_i^2 m_i = 53{,}680.$$

The normal deviate is

$$\frac{+1369.98}{\sqrt{53{,}680}} = +5.91$$

indicating a significant excess of families with even numbers of boys.

The reason why the two subtraction terms are negligible is somewhat peculiar to this example In a binomial with $p = \frac{1}{2}$ and n even, the quantities $\sum g_i m_i$ and $\sum i g_i m_i$ both vanish, as algebraic identities, for this set of g_i. Here we have a binomial with p nearly $\frac{1}{2}$. As an exercise in the computations, the reader may try

$$g_i = +1 \ (i \text{ even}) : g_i = 0 \ (i \text{ odd})$$

so as to test the sum of the even deviations. The normal deviate will again be found to be $+5.91$, as would be expected, but the subtraction terms are not negligible.

Although this example serves to illustrate the computations needed in applying the test to binomial data, there are questions about the validity of the application In working out the frequency distribution of L, we have assumed that the coefficients g_i are chosen before seeing the data, whereas the function L was actually constructed for a type of departure from the binomial that was observed in the data. The effect is to make the L test give too many apparently significant results. This point will be discussed further in section VI. Secondly, as already noted, there is an excess of families with very uneven numbers of boys and girls, so that the binomial model used for the null hypothesis probably does not apply exactly. This disturbance will also influence to some degree the frequency distribution of L, just as non-normality in the basic data influences the t-distribution in a Student's t-test

Example 2. This and example 3 are artificial, and are intended to illustrate two properties of the test.

In 100 families of size 2, the frequency with which some attribute occurs is recorded. The binomial $(q+p)^2$ is fitted, the estimate $\hat{p}$ being 0.2.

In this example the goodness of fit χ^2 has 1 *d.f.* Hence if any one of the single deviations $L=(f_i-m_i)$ is tested by use of equation (16), the χ^2 value for the single deviation should also equal 6 25.

For instance, consider the deviation for $i=2$. From equation (16),

$$\hat{V}(f_2 - m_2) = m_2 - \frac{m_2^2}{N} - \frac{m_2(2-\hat{\mu})^2}{Nn\hat{p}\hat{q}}.$$

Since $\hat{\mu}=n\hat{p}=0\ 4$, this gives

$$\hat{V}(f_2 - m_2) = 4 - \frac{16}{100} - \frac{(16)(2.56)}{(100)(0\ 4)(0\ 8)} = 2.56.$$

Hence

$$\chi_L^2 = \frac{16}{2\ 56} = 6\ 25$$

TABLE 2

GOODNESS OF FIT TEST OF BINOMIAL DISTRIBUTION

i	f_i	m_i	Contr to χ^2
0	60	64	0 25
1	40	32	2 00
2	0	4	4 00
N =	100	100	6 25 (1 *d f*)

The reader may verify that the same value of χ_L^2 is obtained from the deviations for $i=0$ and $i=1$ (Corrections for continuity were omitted in this example, since the purpose is to point out an algebraic relationship)

Example 3 A Poisson distribution is fitted to a sample of size 100, in a situation in which the goodness of fit χ^2 again has 1 degree of freedom (Table 3). If a χ^2 value is computed separately for each of the three deviations, we find $\chi_L^2=0\ 365$ for $i=0$, 0 554 for $i=1$, and 0 728 for $i=2$ Thus the "single deviation" tests give *different* results from one another and from the goodness of fit test, in contrast to the result in Example 2

The discrepancies, which are puzzling at first sight, are a consequence of the grouping of the expectations for $i=2, 3, 4 \cdots$ which occurs in

the 2+ class. On account of this grouping, the value of $\sum im_i$, from Table 3, is 25.743, whereas $\sum if_i$ is 26 Thus the sample value of

$$X = \sum i(f_i - m_i)$$

equals 0.257 instead of 0. The three "single deviation" χ^2 can be brought into much closer agreement with the goodness of fit χ^2 by computing the numerators as

$$[(f_i - m_i) - b_i X]^2,$$

where $X = 0\ 257$

On reflection, however, I doubt whether this adjustment is worthwhile, because the goodness of fit χ^2 is also based on the assumption that $\sum if_i = \sum im_i$ In other words, all four values of χ^2 are approximate, and it is not clear that any one of them should be regarded as superior or preferable.

TABLE 3

GOODNESS OF FIT TEST OF POISSON DISTRIBUTION

i	f_i	m_i	Contr to χ^2
0	78	77.105	0 010
1	18	20 047	0.209
2+	4	2 848	0 466
			0 685 (1 *d f.*)

Example 4 This example illustrates the test of a single deviation in a frequency distribution that is less familiar than the binomial or Poisson. The figures in Table 4 show the number of adult syphilis patients remaining on the roster of a Baltimore clinic at the beginning of successive two-month periods of observation. All the data refer to the same initial group of 232 patients, so that the successive observations are not independent.

The data are suggestive of an exponential decay curve. Perhaps the simplest mathematical framework that might apply is to suppose that in the hypothetical population of which these data are a sample, there is a constant probability p that any person on the roster at the beginning of a two-month period will drop out during the period The proportion of the population dropping out in the ith period is then pq^{i-1}, and the proportion remaining at the end of the 11 periods is q^{11}. On

TABLE 4

NUMBER OF PATIENTS REMAINING ON CLINIC ROSTER AT BEGINNING OF SUCCESSIVE TWO-MONTH PERIODS

Month following admission	Number of patients on roster
0	232
2	171
4	148
6	134
8	121
10	104
12	90
14	78
16	67
18	61
20	56
22	52

this argument, the successive numbers *dropped* in the sample should follow a multinomial distribution with expectations as shown in Table 5. Note that it is necessary to include the number remaining at the end of 11 periods (i e., the number who would drop out in periods 12+) in order that the numbers add to the original total of 232.

TABLE 5

PATIENTS DROPPED FROM ROSTER

Period	Number dropped		$f_i - m_i$	χ^2
	Observed (f_i)	Expected (m_i)		
1 (0–2)	61	33 09	+27 91	23.54
2 (2–4)	23	28 37	− 5 37	1.02
3 (4–6)	14	24 32	−10 32	4 38
4 (6–8)	13	20 85	− 7 85	2 96
5 (8–10)	17	17 88	− 0 88	0 04
6 (10–12)	14	15 33	− 1 33	0 12
7 (12–14)	12	13 14	− 1.14	0.10
8 (14–16)	11	11 27	− 0 27	0.01
9 (16–18)	6	9 66	− 3 66	1 39
10 (18–20)	5	8 28	− 3.28	1 30
11 (20–22)	4	7.10	− 3 10	1 35
12 (22+)	52	42.69	+ 9 31	2.03
	232	231.98		38.24

There is a simple maximum likelihood estimate of p. Let N be the initial number on the roster, N_i the number remaining at the *end* of the ith period, and f_i the number dropped during the ith period, so that $f_1 = N - N_1 : f_i = N_{i-1} - N_i$, $(i = 2, \cdots 11) : f_{12} = N_{11}$.

The probability of the sample is, apart from factors not involving p,

$$P(S) \sim p^{f_1}(pq)^{f_2}(pq^2)^{f_3} \cdots (pq^{10})^{f_{11}} q^{11f_{12}}$$
$$= p^{(f_1 + f_2 + \cdots + f_{11})} q^{(f_2 + 2f_3 + \cdots + 11f_{12})}.$$

But

$$f_1 + f_2 + \cdots + f_{11} = (N - N_1) + (N_1 - N_2) + \cdots + (N_{10} - N_{11})$$
$$= N - N_{11} = D,$$

where D is the total number dropped during the 11 months.

$$f_2 + 2f_3 + \cdots + 11f_{12} = (N_1 - N_2) + 2(N_2 - N_3) + \cdots$$
$$+ 10(N_{10} - N_{11}) + 11N_{11}$$
$$= N_1 + N_2 + \cdots + N_{11} = T - D \text{ (say)}$$

where

$$T = N + N_1 + \cdots + N_{10},$$

is the total of the numbers remaining at the beginning of all periods, excluding the last period. Writing Λ for the log of the likelihood, we have

$$\Lambda = \log P(S) \sim D \log p + (T - D) \log q$$
$$\frac{\partial \Lambda}{\partial p} = \frac{D}{p} - \frac{(T - D)}{q}. \tag{17}$$

This gives

$$\hat{p} = \frac{D}{T}$$

This estimate is a rather natural one The number dropped in any period, divided by the number present at the beginning of the period, is an unbiased estimate of p. The estimate $\hat{p}$ is the total of the numbers dropped, divided by the total of the numbers present.

From Table 4

$$D = 232 - 52 = 180 : T = 232 + 171 + \cdots + 56 = 1262$$
$$\hat{p} = \frac{180}{1262} = 0.14263.$$

The expectations and the deviations are given in Table 5 Clearly, the first period is aberrant, the number dropped being greatly above expectation.

In order to test the deviation for the first period, we use the general formula (14) for its estimated variance, i e ,

$$\hat{V}(f_1 - m_1) \doteq m_1 - \frac{m_1^2}{N} - \frac{1}{I}\left(\frac{\partial m_1}{\partial p}\right)^2 \tag{18}$$

Since $M_1 = Np$,

$$\frac{\partial M_1}{\partial p} = N = 232.$$

In a problem of this type, I is usually most easily found by means of the relation

$$I = E\left(-\frac{\partial^2 \Lambda}{\partial p^2}\right).$$

By differentiating equation (17), we have

$$\frac{\partial^2 \Lambda}{\partial p^2} = \frac{-D}{p^2} - \frac{(T - D)}{q^2} \cdot$$

Since

$$E(D) = N(1 - q^{11}),$$

$$E(T - D) = N(q + q^2 + \cdot \quad + q^{11}) = Nq(1 - q^{11})/p,$$

we find

$$I = E\left(\frac{-\partial^2 \Lambda}{\partial p^2}\right) = \frac{N(1 - q^{11})}{p^2 q}.$$

Substitution in formula (18) gives the desired variance (where the last term is written in a form convenient for computation)

$$\hat{V}(f_1 - m_1) \doteq m_1 - \frac{m_1^2}{N} - \frac{(Np)^2 q}{N(1 - q^{11})}$$

$$= (33\ 09) - \frac{(33\ 09)^2}{232} - \frac{(33\ 09)^2(0\ 85737)}{189\ 31} = 23\ 411.$$

Finally, applying the correction for continuity to $f_1 - m_1 = 27.91$,

$$\chi^2 = \frac{(27.41)^2}{23.411} = 32\ 09$$

As would be anticipated, this value is highly significant.

This deviation can be tested, alternatively, by fitting the same model to the data from the second period onwards, starting with the 171 patients left at the end of the first period. The value of the corresponding goodness of fit χ^2 is 6 18, with 9 *d f.*, as against the original goodness of fit χ^2 of 38 24, with 10 *d f* The difference, 32 06, can be regarded as a test of the fit during the first period The discrepancy between the values 32 06 and 33 27 (the value given by the L test if no correction for continuity is applied) is presumably due to "small sample" effects. Both methods lead to the conclusion that the model fits satisfactorily after the first period, but that the loss during the first period is too large.

VI. TESTS SELECTED AFTER INSPECTION OF THE DATA

In the tests described in this paper, the coefficients g_i must be chosen before inspecting the deviations. With the test for a *single* deviation, it is tempting to apply the test to a deviation for which the contribution to χ^2, i e , $(f_i - m_i)^2/m_i$, looks suspiciously large I should not wish to discourage examination of aberrant individual deviations, but the significance probability given by the L test will then be too low I have not been able to obtain an expression for the significance probability which takes account of the selection after inspection of the deviations From intuitive reasoning, it appears that this probability will lie between P (that given by the L test) and kP, where k is the number of classes in the goodness of fit test In Example 4, there were reasons for suspecting in advance that the first period would show an abnormal drop, since some patients, having learned that they had syphilis, might shrink from the long course of treatment that was then necessary (the data refer to the 1930's), while others might go elsewhere for treatment Thus it might have been decided in advance to test the deviation for the first period On the other hand, if the deviation is picked out purely from inspection, the significance probability appears to lie between P and $12P$. Since P is infinitesimal in this example, there is no doubt about the statistical significance.

The above remarks refer to a single deviation. Suppose now that a linear function L of a number of the deviations is picked out for testing because it looks interestingly large. Then a test that takes account of this selection and errs on the safe side, i.e., gives in general too few significant results, is obtained by referring $L^2/\hat{V}(L)$ to the χ^2 table with the number of degrees of freedom used in the goodness of fit test

To show this, we shall show that if the coefficients g_i in L are selected

so as to make $L^2/\hat{V}(L)$ as large as possible, the maximum value of this quantity is equal to the computed value of χ^2 in the ordinary goodness of fit test. In other words, if we always picked out the linear function that would be the "most significant" of all linear functions, we would obtain a valid test of this function by referring $L^2/\hat{V}(L)$ to the χ^2 table used for the goodness of fit test. Since in practice a linear function that is picked out because it is interesting will not generally give the largest possible normal deviate, we will be making a conservative test of significance if we refer $L^2/\hat{V}(L)$ to the χ^2 table.

In Example 1 the normal deviate was 5 91 Since χ^2 for these data has seven degrees of freedom, a conservative test is to refer $(5.91)^2$ or 34.93 to the χ^2 table with seven degrees of freedom The result is still statistically significant

To prove the needed result, write $\hat{V}$ for $\hat{V}(L)$ Then

$$\frac{\partial}{\partial g_i}\left(\frac{L^2}{\hat{V}}\right) = \frac{2L}{\hat{V}}\frac{\partial L}{\partial g_i} - \frac{L^2}{\hat{V}^2}\frac{\partial \hat{V}}{\partial g_i}.$$

When we set this equal to zero, we obtain the equations

$$\frac{\partial L}{\partial g_i} = \frac{1}{2}\frac{L}{\hat{V}}\frac{\partial \hat{V}}{\partial g_i}. \tag{18a}$$

Now from section III,

$$L = \sum g_i(f_i - m_i) : \qquad \frac{\partial L}{\partial g_i} = (f_i - m_i),$$

$$\hat{V} = \sum g_i^2 m_i - \frac{(\sum g_i m_i)^2}{N} - \frac{1}{I}\left(\sum g_i \frac{\partial m_i}{\partial \theta}\right)^2,$$

$$\frac{1}{2}\frac{\partial \hat{V}}{\partial g_i} = g_i m_i - \frac{m_i}{N}(\sum g_i m_i) - \frac{1}{I}\frac{\partial m_i}{\partial \theta}\left(\sum g_i \frac{\partial m_i}{\partial \theta}\right).$$

Substitution in (18a) gives, on dividing by m_i,

$$\frac{(f_i - m_i)}{m_i} = \frac{L}{\hat{V}}\left\{g_i - \frac{1}{N}(\sum g_i m_i) - \frac{1}{I}\frac{1}{m_i}\frac{\partial m_i}{\partial \theta}\left(\sum g_i \frac{\partial m_i}{\partial \theta}\right)\right\}.$$

Multiply both sides by $(f_i - m_i)$ and add over all classes The left side becomes the computed value of χ^2 in the goodness of fit test. On the right side, the first term becomes

$$\frac{L}{\hat{V}}\left\{\sum g_i(f_i - m_i)\right\} = \frac{L^2}{\hat{V}},$$

as desired to prove the result. The second and third terms on the right both vanish, the second because $\sum f_i = \sum m_i$, the third because of the maximum likelihood equation of estimation for θ. This completes the proof. The result is valid only asymptotically, since the expression for $\hat{V}$ is an approximation. When the expectations are given, it is easy to show that the result holds absolutely. Conceptually, this test is of the same kind as that given by Fisher in §64 of *The Design of Experiments*, and later by Scheffé [5], for testing any linear combination of the treatment means in an analysis of variance.

VII. TWO-PARAMETER ESTIMATION

The "regression" approach extends to the situation where two unknown parameters θ_1 and θ_2 are estimated by maximum likelihood. Since the extension goes smoothly, not all of the details will be presented. We first quote the analogues of the two results (5) and (7) from maximum likelihood theory. The analogue of (5) is the pair of equations

$$\left.\begin{aligned} I_{11}(\hat{\theta}_1 - \theta_1) + I_{12}(\hat{\theta}_2 - \theta_2) &\doteq X_1 \\ I_{12}(\hat{\theta}_1 - \theta_1) + I_{22}(\hat{\theta}_2 - \theta_2) &\doteq X_2 \end{aligned}\right\}, \tag{19}$$

where

$$I_{uv} = \sum \frac{1}{M_i}\left(\frac{\partial M_i}{\partial \theta_u}\right)\left(\frac{\partial M_i}{\partial \theta_v}\right), \qquad u, v = 1, 2$$

$$X_u = \sum \frac{(f_i - M_i)}{M_i}\frac{\partial M_i}{\partial \theta_u}, \qquad u = 1, 2,$$

these being straightforward extensions of the previous notation.

If c_{uv} is the inverse of the matrix I_{uv}, the solutions of these equations are

$$\left.\begin{aligned} \hat{\theta}_1 - \theta_1 &\doteq c_{11}X_1 + c_{12}X_2 \\ \hat{\theta}_2 - \theta_2 &\doteq c_{12}X_1 + c_{22}X_2 \end{aligned}\right\}. \tag{20}$$

The analogue of equation (7) is

$$m_i - M_i \doteq (\hat{\theta}_1 - \theta_1)\frac{\partial M_i}{\partial \theta_1} + (\hat{\theta}_2 - \theta_2)\frac{\partial M_i}{\partial \theta_2}. \tag{21}$$

Hence

$$(f_i - m_i) = (f_i - M_i) - (\hat{\theta}_1 - \theta_1)\frac{\partial M_i}{\partial \theta_1} - (\hat{\theta}_2 - \theta_2)\frac{\partial M_i}{\partial \theta_2}.$$

Substituting from (20) and rearranging, we obtain

$$f_i - m_i \doteq (f_i - M_i) - \left\{c_{11}\frac{\partial M_i}{\partial \theta_1} + c_{12}\frac{\partial M_i}{\partial \theta_2}\right\} X_1$$
$$- \left\{c_{21}\frac{\partial M_i}{\partial \theta_1} + c_{22}\frac{\partial M_i}{\partial \theta_2}\right\} X_2.$$

This is the analogue of the key equation (8) It can be shown that the coefficients of X_1 and X_2 are the regression coefficients of $(f_i - M_i)$ on X_1 and X_2

Hence, as before, the variance of L is equal to the residual variance of L' from its multiple regression on X_1 and X_2 To find this residual variance, we have

$$\text{Cov}\,(L', X_1) = \sum g_i \frac{\partial M_i}{\partial \theta_1} = S_1, \qquad \text{(say)},$$

$$\text{Cov}\,(L', X_2) = \sum g_i \frac{\partial M_i}{\partial \theta_2} = S_2, \qquad \text{(say)}.$$

The two regression coefficients (for L' on X_1, X_2) are

$$b_1 = c_{11}S_1 + c_{12}S_2,$$
$$b_1 = c_{21}S_1 + c_{22}S_2$$

The reduction in variance due to the regression is

$$b_1S_1 + b_2S_2 = c_{11}S_1^2 + 2c_{12}S_1S_2 + c_{22}S_2^2.$$

This gives, finally,

$$V(L) \doteq \sum g_i^2 M_i - \frac{(\sum g_i M_i)^2}{N} - (c_{11}S_1^2 + 2c_{12}S_1S_2 + c_{22}S_2^2)$$

as the general formula for two-parameter estimation

VIII. APPLICATION TO THE NORMAL DISTRIBUTION

In order to obtain a formula applicable to the normal distribution, we may use the fact that the equations of estimation make the mean and variance of the theoretical distribution equal to those of the sample. Let d_i be the center of the ith class It will simplify the algebra if the origin is placed at the sample mean Then we take

$$X_1 = \sum d_i(f_i - M_i),$$
$$X_2 = \sum d_i^2(f_i - M_i).$$

These choices assume that the sample is large and that the grouping into classes is not too coarse. Because of the grouping, $\sum d_i M_i$ and $\sum d_i^2 M_i$ will not exactly equal the first two moments of the continuous normal distribution and it is assumed here that the discrepancies are not serious. Moreover, in practice we equate the sample *mean square* (dividing by $N-1$) to its expectation, instead of the sum of squares, so that terms in $1/N$ are considered negligible

It is convenient to have a general formula for the covariance of any two linear functions

$$H = \sum h_i(f_i - M_i) : K = \sum k_i(f_i - M_i).$$

From equations (2) and (3) it is found that

$$\text{Cov}\ (H, K) = \sum h_i k_i M_i - \frac{(\sum h_i M_i)(\sum k_i M_i)}{N}. \qquad (24)$$

The result remains valid when H and K are identical, in which case the covariance becomes a variance

Hence we obtain the results needed for setting up the regression equations.

$$\text{Cov}\ (L', X_1) = \sum g_i d_i M_i \qquad (25)$$

since $\sum d_i M_i = 0$, because the origin is at the sample mean.

$$\text{Cov}\ (L', X_2) = \sum g_i d_i^2 M_i - \frac{(\sum g_i M_i)(\sum d_i^2 M_i)}{N}$$
$$\doteq \sum g_i M_i (d_i^2 - s^2), \qquad (26)$$

where s^2 is the sample variance.

$$V(X_1) = \sum d_i^2 M_i = N\,s^2, \qquad (27)$$

$$\text{Cov}\ (X_1, X_2) = \sum d_i^3 M_i \doteq 0, \qquad (28)$$

since the third moment of the theoretical normal distribution is zero. Finally,

$$V(X_2) = \sum d_i^4 M_i - \frac{(\sum d_i^2 M_i)^2}{N} \doteq 2Ns^4 \qquad (29)$$

since $\mu_4 = 3\mu_2^2$ for the normal distribution

The five equations (25–29) enable us to set up the regression equations of L' on X_1 and X_2 For the normal distribution, the equations separate, since Cov $(X_1, X_2) = 0$ Hence, the estimated variance is

approximately

$$\hat{V}(L) = \sum g_i^2 m_i - \frac{(\sum g_i m_i)^2}{N} - \frac{[\mathrm{Cov}\,(L, X_1)]^2}{V(X_1)} - \frac{[\mathrm{Cov}\,(L, X_2)]^2}{V(X_2)}$$

$$= \sum g_i^2 m_i - \frac{(\sum g_i m_i)^2}{N} - \frac{(\sum g_i d_i m_i)^2}{Ns^2} - \frac{[\sum g_i m_i (d_i^2 - s^2)]^2}{2Ns^4}$$

APPENDIX

An example of the small-sample distribution of L'

The example refers to a multinomial distribution with three classes The sample size is 10, and the expectations in the three classes are 5, 3 and 2, respectively. This example has been discussed previously, with respect to the ordinary χ^2 test, by Neyman and Pearson [4]. Note that the expectations are fixed. It would be more revealing to work some examples in which the expectations are estimated from the data, but this is considerably more laborious.

The probabilities of each of the 66 possible configurations of the sample were first computed. From these, the exact frequency distributions were worked out for the following two linear functions:

$$L_1' = 6(f_1 - M_1) + 3(f_2 - M_2) + (f_3 - M_3)$$

and

$$L_2' = (f_1 - M_1) + 3(f_2 - M_2) + 6(f_3 - M_3).$$

In L_1', the class with the highest expectation receives the highest weight and the class with the lowest expectation receives the lowest weight In L_2', these weights are reversed. It was thought that the normal approximation might agree better with the exact distribution for L_1' than for L_2'. On the null hypothesis, both L_1' and L_2' have means zero: their standard deviations, as found from equation (4), are 6 3953 and 6 0332, respectively The exact probabilities and the normal approximations for a two-tailed test are shown in Table 6 for the region in which the exact probability lies between 0 25 and 0 005.

For both L_1' and L_2', the normal approximation tends to underestimate the probability, i.e , to give too many apparently significant results Table 6 also shows the errors in the normal probabilities as percentages of the true probabilities. The averages of these percentage errors, ignoring sign, are 12 per cent for L_1' and 10 per cent for L_2'

Since the values of the L''s proceed by integers, it would be easy to apply a correction for continuity in calculating the normal approximation. This correction removes the tendency of the normal approxima-

TABLE 6

COMPARISON OF EXACT PROBABILITIES AND NORMAL APPROXIMATIONS

Deviate	L_1'			L_2'		
	Exact P	Normal approx	Error in %	Exact P	Normal approx.	Error in %
8	.257	.211	−18	.213	.185	−13
9	.172	.159	− 8	.156	136	−13
10	.148	118	−20	.120	0903	−25
11	.0997	.0854	−14	0723	.0683	− 6
12	0656	.0607	− 7	.0563	.0467	−17
13	0549	.0422	−23	0303	.0312	+ 3
14	0289	.0286	− 1	.0210	.0203	− 3
15	0185	.0190	+ 3	.0138	.0129	− 7
16	.0130	0123	− 5	.0084	.0080	− 5
17	.0064	.0079	+23	.0052	.0048	− 8

tion to underestimate the true probabilities. in fact the corrected values are mostly overestimates Apart from this effect, the correction does not improve the approximation. With the correction, the average percentage errors are 15 per cent for L_1' and 12 per cent for L_2'.

Contrary to expectations, the normal approximation is not closer for L_1' than for L_2'. It is, however, closer in single-tailed tests because the distribution of L_1' is not far from symmetrical, whereas that of L_2' is quite skew.

Although the user must judge for himself whether these approximations are satisfactory for practical use, the agreement seems surprisingly good when one considers that the *largest* expectation is 5.

In conclusion, I wish to thank Dr. W Kruskal and Dr P. Meier for some useful suggestions.

REFERENCES

[1] Cochran, W G , "Some methods for strengthening the common χ^2 tests," *Biometrics*, 10 (1954), 417–51

[2] Cramér, H., *Mathematical Methods of Statistics* Princeton· Princeton University Press, 1946

[3] Fisher, R. A , *Statistical Methods for Research Workers*. Edinburgh· Oliver and Boyd (1925), §18 in all editions.

[4] Neyman, J. and Pearson, E. S , "Further notes on the χ^2 distribution," *Biometrika*, 22 (1931), 301–2.

[5] Scheffé, H. "A method for judging all contrasts in the analysis of variance," *Biometrika*, 40 (1953), 87–104.

THE ESTIMATION OF AN OPTIMUM SUBSAMPLING NUMBER*

SAMUEL H. BROOKS
The Johns Hopkins University

I INTRODUCTION

IN TWO-STAGE sampling the usual criterion for the optimum number of subunits to be sampled from a primary unit is that number which minimizes the cost variance product It is well-known that this depends on the ratio of the variance within primary units to the variance between primary units Estimates of this ratio are often made from data at hand or from a pilot sample This paper discusses the estimation of the optimum subsampling number, m_{op}, such that when this estimate of m_{op} is used to take the main sample, the precision of the resulting estimate of the population mean averages at least 90 per cent of the precision obtainable if m_{op} were used It is shown that in some cases an adequate subsampling number can be estimated without a pilot sample

II ELEMENTARY THEORY

This section, containing some of the elementary definitions and results of two-stage sampling theory, is included in order to have a unified presentation. More complete discussions can be found in Deming [3], Cochran [2], and Hansen, Hurwitz and Madow [5]

A The subsampling number in two-stage sampling

Two-stage sampling is often used to estimate the mean of a population comprised of units which contain the fundamental elements of the population. The first stage is the selection of a sample of units The second stage is the subsampling of these units The subsampling number is the number of elements sampled within each of the units selected in the first stage. These units and elements are frequently called primary units and secondary units respectively.

An agricultural study of the sugar content of a variety of sugar beets may serve as an example of two-stage sampling The statistic of interest is the average per cent sugar content of the beets in a field. The

* Department Paper No 294, Department of Biostatistics, The Johns Hopkins School of Hygiene and Public Health Most of the work in this paper was done under Navy Contract Nonr—248 (16) and was completed under a Johns Hopkins University fellowship

field is divided into a number of plots. The first stage of sampling is the selection of n plots to be sampled. The second stage is the sampling of m beets from each plot selected.

Two-stage sampling is to be distinguished from sequential procedures such as two-sample sampling and double sampling in which the results of the first sample indicate how a second sample should be made However, it is clear that two-stage sampling may be used in a two-sample procedure if the population has the appropriate structure. In fact, the pilot sample designs in this paper may be considered the first sample of a two-sample procedure

Precision and economy are the criteria in the choice of an appropriate subsampling number. If the elements vary little within a unit and vary much from unit to unit, it is reasonable to use a low subsampling number However, if the sampling of a unit is very expensive relative to the sampling of an element, a large subsampling number is appropriate

B The model of the population and of the cost of a sample

$\overline{Y}$ = population mean
u_i = deviation of the mean of the ith unit from the population mean
e_{ij} = deviation of the jth element of the ith unit from the mean of the ith unit
y_{ij} = value of the jth element of the ith unit
M = number of elements contained in each unit, *constant for all units in the population*
N = number of units in the population

$$y_{ij} = \overline{Y} + u_i + e_{ij}. \tag{1}$$

The average value of u_i is zero, and within the ith unit the average value of e_{ij} is zero The variance of u_i is S_u^2, and the variance of e_{ij} is S_w^2. Analogous to their meanings for infinite populations these variances are defined for finite populations by the relations:

$$S_w^2 = \frac{1}{(M-1)N}\sum_{i=1}^{N}\sum_{j=1}^{M}(y_{ij} - \overline{Y}_i)^2,$$

$$S_w^2 + MS_u^2 = \frac{M}{N-1}\sum_{i=1}^{N}(\overline{Y}_i - \overline{Y})^2,$$

where $\overline{Y}_i$ is the mean value of the ith unit

Both units and elements within units are selected by random sampling, so that the variates u_i and e_{ij} are independently distributed.

C_u = cost of selecting a single unit for subsampling
n = number of units selected for subsampling
C_e = cost of sampling an element within a selected unit
m = number of elements sampled within each selected unit (subsampling number)
C_s = cost of executing the two-stage sample

$$C_s = C_u n + C_e nm. \tag{2}$$

C Sample mean and variance

An unbiased estimate of the population mean per element is the simple average of all the elements in the sample,

$$\bar{y} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij}$$

The variance of this estimate is

$$V(\bar{y}) = \left(\frac{1}{n} - \frac{1}{N}\right) S_u^2 + \left(\frac{1}{nm} - \frac{1}{NM}\right) S_w^2. \tag{3}$$

An unbiased estimate of $V(\bar{y})$ from the sample is

$$v(\bar{y}) = \frac{1}{m}\left(\frac{1}{n} - \frac{1}{N}\right) s_b^2 + \frac{1}{N}\left(\frac{1}{m} - \frac{1}{M}\right) s_w^2$$

where s_b^2 is the between units mean square and s_w^2 is the within units mean square from the sample as computed in an analysis of variance table

D The optimum subsampling number

The value of m is here found which will minimize the variance of the sample estimate of the population mean when the cost of the two-stage sample is fixed From equations (2) and (3),

$$\frac{1}{n} = \frac{C_e}{C_s}\left(\frac{C_u}{C_e} + m\right),$$

$$V(\bar{y}) = \frac{S_u^2}{n}\left(1 + \frac{1}{m}\frac{S_w^2}{S_u^2}\right) - \frac{S_u^2}{N}\left(1 + \frac{1}{M}\frac{S_w^2}{S_u^2}\right).$$

Eliminating $1/n$,

$$V(\bar{y}) = \frac{C_e S_u^2}{C_s}\left(\frac{C_u}{C_e} + \frac{S_w^2}{S_u^2} + m + \frac{1}{m}\frac{C_u}{C_e}\frac{S_w^2}{S_u^2}\right) - \frac{S_u^2}{N}\left(1 + \frac{1}{M}\frac{S_w^2}{S_u^2}\right). \tag{4}$$

By setting the derivative of $V(\bar{y})$ with respect to m equal to zero, m_{op}, the value of m which minimizes the variance of the estimate of the mean when the sample cost is fixed, is found:

$$\frac{dV(\bar{y})}{dm} = \frac{C_e S_u^2}{C_s}\left(1 - \frac{1}{m^2}\frac{C_u}{C_e}\frac{S_w^2}{S_u^2}\right) = 0.$$

The optimum subsampling number is

$$m_{op} = \sqrt{\frac{C_u}{C_e}}\,\frac{S_w}{S_u} \tag{5}$$

If $V(\bar{y})$ is fixed, equation (4) may be solved for C_s,

$$C_s = \frac{C_e}{\dfrac{V(\bar{y})}{S_u^2} + \dfrac{1}{N}\left(1 + \dfrac{1}{M}\dfrac{S_w^2}{S_u^2}\right)}\left(\frac{C_u}{C_e} + \frac{S_w^2}{S_u^2} + m + \frac{1}{m}\frac{C_u}{C_e}\frac{S_w^2}{S_u^2}\right) \tag{6}$$

It may be seen that the variable m appears in equation (6) in the same manner as in equation (4), so that the m_{op} of equation (5) is also the value of m which will minimize the cost of the sample if the variance of the estimate of the mean is fixed, thus it minimizes the cost variance product. The quantitative expression for m_{op} in equation (5) is consistent with the criteria in section II A

The corresponding number of units to be sampled, n, is obtained from equation (2) if the cost of the sample is fixed, or from equation (3) if the variance of the estimate of the mean is fixed.

If M, N, C_u, C_e, S_u^2, and S_w^2, are well known, and m_{op} is used, then the cost of a two-stage sample which will yield an estimate of the mean having the desired or target variance $T(\bar{y})$ is

$$C_s = \frac{(\sqrt{C_u}\,S_u + \sqrt{C_e}\,S_w)^2}{T(\bar{y}) + \dfrac{1}{N}\left(S_u^2 + \dfrac{1}{M}S_w^2\right)}\,.$$

E. Estimation of m_{op}

The optimum subsampling number may be estimated from a pilot sample or from an available array of data which may be considered as a pilot sample. Let h be the number of units selected for sampling in the first stage of the pilot sample, and let k be the subsampling number used in the second stage of the pilot sample The numbers h and k will be used as the dimensions of the pilot sample, and the numbers n and m will be the corresponding dimensions of the subsequent main sample. It is assumed that the cost ratio C_u/C_e is known exactly That this assumption is not critical will be shown later Thus, the estimate of m_{op} is dependent simply on the estimate of the variance ratio, S_w^2/S_u^2 which is obtained from the analysis of variance table resulting from the pilot sample:

Source of Variation	Degrees of Freedom	Mean Squares from Sample	Expected Value of Mean Squares
Between units	$h-1$	s_b^2	$S_b^2=S_w^2+kS_u^2$
Within units between elements	$h(k-1)$	s_w^2	S_w^2

It may be seen from the expected values that

$$\frac{S_w}{S_u}=\left\{\frac{S_w^2}{\frac{1}{k}(S_b^2-S_w^2)}\right\}^{1/2}=\left\{\frac{k}{\frac{S_b^2}{S_w^2}-1}\right\}^{1/2}.$$

So that the estimate of S_w/S_u from the sample is

$$\frac{s_w}{s_u}=\left[\frac{k}{\frac{s_b^2}{s_w^2}-1}\right]^{1/2}.$$

The estimate of the optimum subsampling number, which shall be denoted by $\widehat{m}_{op}$, is obtained from equation (5) by the direct substitution of the above estimate,

$$\widehat{m}_{op}=\left[\frac{k\dfrac{C_u}{C_e}}{\dfrac{s_b^2}{s_w^2}-1}\right]^{1/2}. \tag{7}$$

Ordinarily, an integral value of $\widehat{m}_{op}$ is required. The estimate, formula (7), does not yield such a value in general. The procedure for selecting the subsampling number in this case is to use the integer m which satisfies the relation[1]

$$m(m-1) \leqq \widehat{m}_{op}^2 < m(m+1)$$

As an example, suppose that $\widehat{m}_{op}=7.50$ has been obtained from equation (7) Since $8(7) \leqq (7\ 50)^2 < 8(9)$, it may be seen that the subsampling number to be used is $m=8$. It is of interest to note that $m=\infty$ is indicated whenever $s_b^2/s_w^2 \leqq 1$, implying that each unit in the sample should have all its elements enumerated

It is conceivable that a non-integral value might be used for the subsampling number This effect may be achieved by using one subsampling number on some proportion of the units and another subsampling number on the rest of the units which are in the sample. However, such efforts would result in a negligible gain in precision which would be offset by the necessary bookkeeping.

III THE EXPECTED RELATIVE PRECISION OF A PILOT SAMPLE DESIGN

A *The relative precision of a subsampling number*

We shall define the relative precision, RP, of a subsampling number m as the ratio of the variance of $\bar{\bar{y}}$ given by m_{op} to the variance of $\bar{\bar{y}}$ given by m for the same sample cost. If the finite population correction can be ignored—that is, if N is large relative to n—then from equation (4)

$$RP(m) = \frac{\dfrac{C_eS_u^2}{C_e}\left(\dfrac{C_u}{C_e} + \dfrac{S_w^2}{S_u^2} + m_{op} + \dfrac{1}{m_{op}}\dfrac{C_u}{C_e}\dfrac{S_w^2}{S_u^2}\right)}{\dfrac{C_eS_u^2}{C_e}\left(\dfrac{C_u}{C_e} + \dfrac{S_w^2}{S_u^2} + m + \dfrac{1}{m}\dfrac{C_u}{C_e}\dfrac{S_w^2}{S_u^2}\right)}.$$

Substitution for m_{op} from equation (5) gives

$$RP(m) = \frac{\dfrac{C_u}{C_e} + \dfrac{S_w^2}{S_u^2} + 2\sqrt{\dfrac{C_u}{C_e}}\dfrac{S_w}{S_u}}{\dfrac{C_u}{C_e} + \dfrac{S_w^2}{S_u^2} + m + \dfrac{1}{m}\dfrac{C_u}{C_e}\dfrac{S_w^2}{S_u^2}}. \tag{8}$$

[1] Cameron [1] indicates that this procedure was originated by Churchill Eisenhart.

B. Distribution of the estimate of the optimum subsampling number

If u_i and e_{ij} may be considered normally distributed, the distribution of s_w^2 is $S_w^2\chi^2$ with $h(k-1)$ degrees of freedom and s_b^2 is distributed as $(S_w^2+kS_u^2)\chi^2$ with $(h-1)$ degrees of freedom. The ratio

$$\frac{\dfrac{s_w^2}{S_w^2}}{\dfrac{s_b^2}{S_w^2+kS_u^2}}=\frac{1+k\dfrac{S_u^2}{S_w^2}}{\dfrac{s_b^2}{s_w^2}}$$

is distributed as F where F has $h(k-1)$ and $(h-1)$ degrees of freedom. From equation (7)

$$\frac{s_b^2}{s_w^2}=1+\frac{k}{\widehat{m}_{op}^2}\frac{C_u}{C_e}.$$

By substitution the random variable $\widehat{m}_{op}$ can be related to F,

$$F=\frac{1+k\dfrac{S_u^2}{S_w^2}}{1+\dfrac{k}{\widehat{m}_{op}^2}\dfrac{C_u}{C_e}}$$

For computational convenience this may expressed in terms of the incomplete β-function, using Karl Pearson's notation The parameters are

$$p=\frac{h(k-1)}{2},$$

$$q=\frac{h-1}{2}$$

The random variable is

$$x(\widehat{m}_{op})=\frac{pF}{pF+q}$$

C Expected relative precision of a pilot sample design

We shall consider a pilot sample design to be the specification of h, the number of units to be selected for sampling, and of k, the subsampling number The units and elements are to be selected randomly

in the pilot sample as well as in the main sample. For the resultant main sample design the subsampling number m is obtained from the pilot sample as indicated in Section II E, and n, the number of units to be selected for subsampling, is found from equation (2) since it is assumed that the cost of the main sample has been fixed. We shall define the expected relative precision of a pilot sample design as being the average relative precision of m's which result from the use of that design for fixed cost and variance ratios.

The relative frequency, $f(m)$, with which the integer m is indicated as the subsampling number by a pilot sample design is the relative frequency that the design yields an estimate of the subsampling number, $\hat{m}_{op}$, equation (7), which is between $\sqrt{m(m-1)}$ and $\sqrt{m(m+1)}$,

$$f(m) = \frac{1}{B(p, q)} \int_{x(\sqrt{m(m-1)})}^{x(\sqrt{m(m+1)})} x^{p-1}(1 - x)^{q-1} dx.$$

Since this m has the relative precision given by equation (8), the expected relative precision, ERP, of the pilot sample design is

$$ERP = \sum_{m=1}^{M} RP\,(m) f(m) \tag{9}$$

When

$$\frac{C_u}{C_e} \cdot \frac{S_w^2}{S_u^2}$$

is greater than 16 and M is large, this summation is well approximated by the integral

$$ERP = \frac{1}{B(p, q)} \int_0^1 RP(m(x)) x^{p-1}(1 - x)^{q-1} dx \tag{10}$$

where

$$m(x) = \left[\frac{k \dfrac{C_u}{C_e}}{\dfrac{p}{q}\left(\dfrac{1 - x}{x}\right)\left(1 + k \dfrac{S_u^2}{S_w^2}\right) - 1} \right]^{1/2}$$

It may be seen that the expected relative precision of a pilot sample design is determined by h, k, M, C_u/C_e and S_w^2/S_u^2.

IV. PILOT SAMPLE DESIGNS HAVING AN EXPECTED RELATIVE PRECISION OF 90 PER CENT

A. *Construction of Table I*

The pilot sample designs of Table I were found by using equations (9) and (10) For fixed cost and variance ratios and $M=\infty$, the expected relative precisions of combinations of h and k were computed. For each h considered, the value of k for which $ERP=90$ per cent was found by inverse interpolation Of these, the combination of h and k which forms the least expensive pilot sample was taken as the "optimum" design.

Preliminary work had indicated that the expected relative precision of a pilot sample design was relatively insensitive to the cost ratio. Therefore, the values of C_u/C_e which were considered in the calculations were rather widely spaced They were 0 01, 1 0, 16, and 100.

The variance ratios considered in the calculations were $S_w^2/S_u^2=0$ 25, 1 0, 4 0, 16, and 64. It is clear from Section V that when the variance ratio is known to be small a subsampling number with high relative precision can be determined without resorting to a pilot sample Pilot sample designs corresponding to variance ratios of 0 5 or less were included in Table I, but the reader is advised to consider Section V carefully if he feels that the variance ratio of the population he intends to study is as small as this

The pilot sample designs of Table I were interpolated from the designs computed for the cost and variance ratios indicated above. Only one value of the parameter M, the number of elements contained in each unit, was considered The pilot sample designs of Table I correspond to the value $M=\infty$

For each combination of the above parameters, equation (9) or (10) was used to determine the expected relative precision of a number of pilot sample designs For each h considered, the ERP was computed for several values of k By graphical interpolation the value of k was found which corresponded to $ERP=90$ per cent. By this means, several pilot sample designs having $ERP=90$ per cent were found for each combination of the cost and variance ratios.

It was then pertinent to consider which of these pilot sample designs would cost the least to execute. The assumption was made that the cost ratio for the pilot sample is the same as for the main sample This does not imply that the per unit and per element costs in the pilot sample must be the same as those in the main sample, but does imply a proportionality between these costs. If this constant of propor-

TABLE I

PILOT SAMPLE DESIGNS HAVING AN EXPECTED RELATIVE PRECISION OF 90%

The number of units to be selected for sampling is h The subsampling number to be used is k The total number of elements to be sampled is hk The cost ratio, C_u/C_e, is the cost of selecting a single unit for subsampling relative to the cost of sampling a single element within a selected unit. The variance ratio, S_w^2/S_u^2, is the "within units" variance relative to the "between units" component of variance

C_u/C_e	≦1			2			4			8			16			32			64			100		
S_w^2/S_u^2	h	k	hk	h	k	hk	h	k	hk	h	k	hk	h	k	hk	h	k	hk	h	k	hk	h	k	hk
≦½	5	3	15	5	3	15	5	3	15	4	4	16	4	4	16	3	7	21	3	8	24	3	9	27
1	7	3	21	6	4	24	6	5	30	5	6	30	5	7	35	4	10	40	4	12	48	3	18	54
2	8	5	40	7	7	49	6	9	54	6	9	54	5	13	65	5	14	70	4	20	80	3	31	93
3	9	7	63	8	8	64	7	10	70	7	11	77	6	15	90	5	20	100	4	28	112	4	31	124
4	9	9	81	8	11	88	8	12	96	7	14	98	7	15	105	5	25	125	5	27	135	4	38	152
6	10	11	110	9	13	117	8	16	128	8	17	136	7	21	147	6	27	162	5	37	185	4	50	200
8	10	14	140	10	15	150	9	17	153	9	18	162	8	22	176	6	32	192	5	44	220	4	60	240
12	10	20	200	10	20	200	10	21	210	9	25	225	8	29	232	7	37	259	5	58	290	5	62	310
16	10	25	250	10	27	270	10	27	270	10	28	280	8	37	296	7	46	322	6	60	360	5	78	390
24	10	35	350	10	36	360	10	37	370	10	38	380	9	45	405	7	62	434	6	80	480	5	104	520
32	10	46	460	10	47	470	10	48	480	10	49	490	9	58	522	8	69	552	6	102	612	5	132	660
48	10	69	690	10	70	700	10	71	710	10	73	730	10	77	770	8	102	816	7	129	903	6	159	954
64	10	92	920	10	93	930	10	96	960	10	100	1,000	10	104	1,040	8	137	1,096	7	169	1,183	6	210	1,260

tionality is r then, analogous to equation (2), the cost of a pilot sample is

$$C_p = rC_u h + rC_e hk.$$

For comparison purposes it is sufficient to consider a pilot sample cost function, C_q, which is proportionate to C_p,

$$C_q = \frac{C_p}{rC_e} = h\left[\frac{C_u}{C_e} + k\right]. \tag{11}$$

By using equation (11) the relative costs of executing the pilot sample designs can be compared As an example, consider the designs that were found to have $ERP = 90$ per cent for $C_u/C_e = 16$ and $S_w^2/S_u^2 = 64$,

h	k	C_q
5	333	1,745
7	171	1,309
9	118	1,206
11	93	1,199
13	78	1,222
15	68	1,260

C_q is plotted against h. C_q has a minimum value of 1,198 near $h = 10$. The corresponding value of k is found by substituting these values in equation (11),

$$1198 \fallingdotseq 10(16 + k)$$

The integer $k = 104$ satisfies this relation. Therefore, the "optimum" sample design in this case is $h = 10$, $k = 104$

B. Application of Table I

Table I contains designs of pilot samples to be used to estimate the optimum subsampling number such that, when this estimate is used in the main sample, it will have a relative precision of 90 per cent on the average.

A population for which the designs of Table I are applicable fits the model described in Section II B. It is to be comprised of an infinite number of units each containing an infinite number of elements, that is, $N = M = \infty$. Furthermore, u_i and e_{ij} are to be normally distributed. If the population does not fit the assumption that N and M both be very large, then the designs of Table I are conservative in the sense that when applied to such a population, these designs would have an

expected relative precision of greater than 90 per cent. It is also assumed that the cost ratio is the same for the pilot sample as that for the main sample, but moderate deviation from this assumption has little effect.

The fundamental assumption of an equal number of elements in each unit should be carefully considered. Failure to have units of nearly equal size, say within 5 per cent of their average, may cause a greater loss of precision. In many cases this difficulty may be alleviated by carefully redefining the unit. For example, the block is a convenient unit in sampling the households of a city for some characteristic. However, in sparsely populated regions of the city several blocks may be treated as one unit, and in densely populated regions a part of a city block is considered to be the unit. By this means units of approximately equal size might be constructed

A cost ratio and a variance ratio must be estimated in order to select a pilot sample design from Table I. The cost ratio may be obtained from the experience of previous investigations or might be estimated from a knowledge of the nature of the proposed investigation. Even a rather crude approximation to the cost ratio could be used without much reduction in the expected relative precision. The value of the variance ratio to be used in entering Table I is the highest such ratio believed possible in the population to be sampled Thus, if this extreme estimate is in fact the "true" variance ratio of the population, the pilot sample design will have an expected relative precision of 90 per cent. If this estimate is greater than the "true" variance ratio, the expected relative precision is greater than 90 per cent

C An example of the use of Table I

In the manufacture of printed cardboard cartons such as breakfast cereal boxes, the moisture content of the cardboard is an important factor affecting the alignment of the printed colors. Therefore, one measure of the quality of a shipment of cardboard received from a supplier by a printing concern is the average moisture content of the cardboard in that shipment. One such concern, the Lord Baltimore Press in Baltimore, Maryland, has recently developed a sampling procedure to make such determinations.

A shipment of cardboard is well suited to two-stage sampling. The cardboard is delivered in 2,000 sheet packages called skids. A single shipment may consist of 300 skids. A skid may be regarded as a unit and the measurements of the moisture content of the sheets within a skid may be regarded as the elements of that unit.

In order to sample from this type of population, an estimate of the

optimum subsampling number is needed. To make this estimate, a pilot sample design was selected from Table I on the basis of estimates of the cost ratio and of an upper bound to the possible values of the variance ratio. It takes about six minutes to move and open a skid and about three minutes to make a moisture content determination, so that an estimate of the cost ratio, C_u/C_e, is $6/3=2$. It was not believed that the variation of moisture content within a skid would be more than twice the variation of moisture content from skid to skid, so that an estimate of an upper bound to the variance ratio, S_w^2/S_u^2, is 2. From these parameters, Table I indicates that the appropriate pilot sample should consist of $k=7$ moisture content determinations to be made at random from each of $h=7$ skids sampled at random from the shipment. The results of the pilot sample are presented in Table II.

TABLE II

MEASUREMENTS OF PERCENTAGE MOISTURE CONTENT IN A SHIPMENT OF 300 SKIDS

Within Skid Measurement Number	Skid Number (Coded)							
	27	102	116	231	235	285	293	
1	6.5	6.3	6.8	6.1	6.7	6.0	6.5	
2	6.0	6.5	7.2	5.8	6.5	5.8	6.8	
3	6.8	6.4	6.8	6.7	6.6	6.2	6.7	
4	6.5	6.5	7.1	6.5	6.5	6.0	7.0	
5	6.3	6.4	7.0	6.0	6.6	6.1	7.1	
6	6.4	6.6	6.7	6.3	6.3	6.1	6.9	
7	6.9	6.3	7.1	5.9	6.5	6.2	6.7	Grand Total
Skid Totals	45.4	45.0	48.7	43.3	45.7	42.4	47.7	318.2

The analysis of variance is computed in the usual manner:

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares
Between Skids	6	4.2589	$0.7098=s_b^2$
Between Measurements within Skids	42	1.9486	$0.0464=s_w^2$
Total	48	6.2075	

From equation (7) the estimate of the optimum subsampling number is

$$\hat{m}_{op} = \left[\frac{(7)(2)}{\dfrac{0.7098}{0.0464} - 1}\right]^{1/2} = 0.990.$$

Since $1(1-1) \leqq (0.990)^2 < 1(1+1)$, it may be seen from Section II E that the subsampling number to be used is $m=1$.

V THE RANGE OF VARIANCE RATIOS FOR WHICH A PARTICULAR SUBSAMPLING NUMBER IS ACCEPTABLE

A particular subsampling number, m_o, having been selected, it is of interest to know for what interval of variance ratios m_o will have an acceptable relative precision when the cost ratio is fixed. We shall designate the least acceptable $RP(m_o)$ as L The limits of this interval of variance ratios is found from equation (8) by solving for S_w^2 S_u^2 in terms of C_u/C_e, L, and m_o,

$$\frac{S_w^2}{S_u^2} = \left[\frac{\sqrt{\dfrac{C_u}{C_e}} \pm \left[m_0 + \dfrac{C_u}{C_e}\right]\sqrt{\dfrac{L - L^2}{m_0}}}{\dfrac{L}{m_0}\dfrac{C_u}{C_e} + L - 1}\right]^2.$$

When the least acceptable relative precision of m_o is 90 per cent, that is, $L = 90$, these limits are given by

$$\frac{S_w^2}{S_u^2} = \left[\frac{10\sqrt{\dfrac{C_u}{C_e}} \pm \dfrac{3}{\sqrt{m_0}}\left[m_0 + \dfrac{C_u}{C_e}\right]}{\dfrac{9}{m_0}\dfrac{C_u}{C_e} - 1}\right]^2. \tag{12}$$

Table III was computed from equation (12). For several cost ratios it indicates the interval of the variance ratio over which the subsampling number tabulated will have a relative precision of at least 90 per cent. In terms of the example in the previous section in which the cost ratio is $C_u/C_e = 2$ and the subsampling number selected is $m_o = 1$, Table III shows that this subsampling number will have a relative precision of at least 90 per cent if, in fact, the true variance ratio is between 0 0 and 1 9 The estimate of the variance ratio from the pilot sample is 0 49

It is possible to find a moderate subsampling number which has

TABLE III

LIMITS DEFINING THE INTERVAL OF THE VARIANCE RATIO OVER WHICH THE SUBSAMPLING NUMBER m_0 WILL HAVE A RELATIVE PRECISION OF AT LEAST 90%

The cost ratio, C_u/C_e, is the cost of selecting a single unit for subsampling relative to the cost of sampling a single element within a selected unit. The variance ratio, S_w^2/S_u^2, is the "within units" variance relative to the "between units" component of variance.

C_u/C_e	1		2		4		8		16		32		64		100	
Limit m_0	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
1	0.0	4.0	0.0	1.9	0.0	1.0	0.0	0.6	0.0	4.0	0.0	0.3	0.0	0.2	0.0	0.2
2	1.1	21.9	0.5	8.0	0.2	3.7	0.0	2.0	0.0	1.2	0.0	0.8	0.0	0.6	0.0	0.5
3	2.4	71.6	1.2	20.8	0.5	8.5	0.2	4.2	0.0	2.4	0.0	1.5	0.0	1.1	0.0	0.9
4	4.0	196.0	2.2	43.7	1.0	16.0	0.4	7.4	0.1	4.0	0.0	2.4	0.0	1.6	0.0	1.3
5	5.9	509.1	3.3	81.9	1.6	26.8	0.7	11.6	0.2	6.0	0.0	3.5	0.0	2.3	0.0	1.8
6	8.1	1,379.9	4.7	143.3	2.4	41.6	1.0	17.1	0.3	8.5	0.1	4.8	0.0	3.0	0.0	2.4
7	10.6	4,455.6	6.3	240.1	3.3	61.4	1.5	23.8	0.5	11.4	0.1	6.3	0.0	3.9	0.0	3.0
8	13.2	24,450.8	8.0	392.0	4.3	87.4	2.0	32.0	0.7	14.8	0.2	8.0	0.0	4.9	0.0	3.7
9	16.0	∞	9.9	632.1	5.4	121.0	2.6	41.9	1.0	18.8	0.3	9.9	0.0	5.9	0.0	4.5
10	19.0	∞	11.9	1,018.1	6.7	163.9	3.3	53.5	1.3	23.3	0.4	12.0	0.0	7.0	0.0	5.2
12	25.3	∞	16.3	2,759.7	9.4	286.6	4.8	83.2	2.1	34.1	0.6	16.9	0.1	9.6	0.0	7.1
16	39.5	∞	26.4	48,901.6	16.0	784.0	8.6	174.9	4.0	64.0	1.5	29.7	0.3	16.0	0.1	11.5
25	76.6	∞	54.0	∞	34.9	7,225.0	20.4	654.2	10.5	184.2	4.5	74.5	1.5	36.6	0.5	25.0
50	201.4	∞	153.1	∞	108.0	∞	69.8	14,450.0	40.7	1,308.3	20.9	368.4	9.0	148.9	4.6	92.7
100	497.6	∞	402.8	∞	306.3	∞	261.1	∞	139.7	777,603.2	81.5	2,616.7	41.9	736.7	25.0	400.0

good relative precision for high variance ratios. The least subsampling number which will have a relative precision L for an infinite variance ratio is

$$m = \left[\frac{L}{(1 - L)}\right]\frac{C_u}{C_s}.$$

This subsampling number also has relative precision L for variance ratios as small as

$$\left[\frac{(L - 0.5)}{(1 - L)}\right]^2 \frac{C_u}{C_s}.$$

For example, when $L = .90$, the subsampling number $9C_u/C_s$ will have a relative precision of at least 90 per cent for all variance ratios greater than $16C_u/C_s$. By this means a subsampling number having acceptable relative precision can be selected when it is known that the variance ratio is high.

It is also possible to find a subsampling number having good relative precision over a large interval of small variance ratios The greatest subsampling number having relative precision L for a zero variance ratio is

$$m = \left[\frac{(1 - L)}{L}\right]\frac{C_u}{C_s}$$

if this quantity is greater than one In this case, this subsampling number also has relative precision L for variance ratios as large as

$$\left[\frac{(1 - L)}{(L - 0.5)}\right]^2 \frac{C_u}{C_s}$$

For example, when $L = .90$ and $(C_u/C_s) \geq 9$, the subsampling number

$$m = \left[\frac{1}{9}\right]\frac{C_u}{C_s}$$

will have a relative precision of at least 90 per cent for all variance ratios less than

$$\left[\frac{1}{16}\right]\frac{C_u}{C_s}.$$

If

$$\left[\frac{(1 - L)}{(L)}\right]\frac{C_u}{C_s}$$

is less than one, the subsampling number $m=1$ may be considered to have at least relative precision L for the zero variance ratio, since this is the least subsampling number possible. Thus, a subsampling number having acceptable relative precision can be selected when it is known that the variance ratio is low.

It may be inferred from this discussion and from Table III that a subsampling number having high relative precision may be estimated without a pilot sample if the order of magnitude of the variance ratio is known, or if the variance ratio is known to be quite large or quite small.

VI OTHER PROBLEMS CONCERNING THE USE OF A PILOT SAMPLE IN THE DESIGN OF A TWO-STAGE SAMPLE

This paper has indicated methods of designing two-stage samples in several situations. When the cost and variance parameters are well known, the results of the elementary theory reviewed in Section II may be used. When the cost of the main sample has been fixed and the cost and variance ratios are known roughly, the results of Section V indicate the adequacy of a subsampling number that has been inferred without benefit of a pilot sample. When the cost of the main sample has been fixed and an upper limit to the variance ratio may be asserted, a pilot sample design can be selected from Table I to be used to estimate the optimum subsampling number. When a subsampling number has been determined and the cost of the main sample has been fixed, the number of units to be sampled is found from equation (2).

There are other situations, however, in which the investigator would like to obtain an estimate of the population mean having a specified variance, this to be done at least cost. Here the pilot sample would be used to estimate the variance components themselves, not just their ratio, and these variance components are used to estimate the number of units to be sampled as well as the subsampling number. The adequacy of such a design indicated by the pilot sample may be judged in terms of how near the actual variance of this design is to the desired variance and how low the cost of executing this design is.

The pilot sample may also be used to estimate the population mean and to furnish information as to how large an additional sample need be to provide a combined estimate of the population mean having a specified variance. In this situation pilot sample designs that are optimum in any reasonable sense seem quite difficult to construct.

VII. ACKNOWLEDGMENTS

I am indebted to William G Cochran for suggesting the problem and for his valuable advice in its solution. Most of the calculations were made at the Johns Hopkins Computing Laboratory, the facilities of which were made available to me by Wade M Etchison The data of the example was provided by Walton Handcock of the Lord Baltimore Press I am also grateful to the members of the Department of Biostatistics for their many helpful suggestions and criticisms

REFERENCES

[1] Cameron, J. M, "Use of variance components in preparing schedules for the sampling of baled wool," *Biometrics*, 7 (1951), 83–97.

[2] Cochran, W G, *Sampling Techniques.* New York John Wiley and Sons; London. Chapman and Hall (1953), Chapter 10.

[3] Deming, W E, *Some Theory of Sampling* New York John Wiley and Sons; London Chapman and Hall (1950), Chapter 5.

[4] Eisenhart, C, "The assumptions underlying the analysis of variance," *Biometrics*, 3 (1947), 18.

[5] Hansen, M H, Hurwitz, W N, and Madow, W. G, *Sample Survey Methods and Theory.* New York John Wiley and Sons; London: Chapman and Hall (1953), Volume 1, Chapter 6, Part E

[6] Immer, F R, "A study of sampling technic with sugar beets," *Journal of Agricultural Research,* 44 (1952), Number 8

[7] Pearson, K, *Tables of the Incomplete Beta-Function* London: Biometrica Office, 1934.

[8] Soper, H E, *The Numerical Evaluation of the Incomplete β-Function* London Cambridge University Press, 1921.

[9] Yates, F, *Sampling Methods for Censuses and Surveys* New York Hafner Publishing Company (1949), Chapter 8 "Efficiency."

EXTENDED TABLES FOR USE WITH THE "G" TEST FOR MEANS

J. EDWARD JACKSON AND ELEANOR L. ROSS
Eastman Kodak Company

This paper uses existing statistical theory to transform the tables of Lord into a form such that the significance tests may be made quickly without the aid of either a desk calculator or a slide rule.

INTRODUCTION

THERE has been a trend in the past few years towards the use of rapid approximate statistical methods. These methods are often somewhat inefficient but this loss in efficiency is usually more than compensated by the amount of time and effort saved in computing the results Such a case is the "G" test for means as an approximation to the one-sample "t" test of the form

$$t = \frac{|\overline{X} - \mu|}{\dfrac{s}{\sqrt{n}}},$$

where

$$\overline{X} = \text{Sample Mean}$$

$$\mu = \text{Hypothetical Mean}$$

$$s = \sqrt{\frac{n\Sigma X^2 - (\Sigma X)}{n(n-1)}}$$

$$n = \text{Sample Size.}$$

The alternative "G" test replaces $s/\sqrt{n}$ by R, the sample range, and is of the form

$$G = \frac{|\overline{X} - \mu|}{R}.$$

The distribution of this statistic was found by Daly [1] in 1946

Lord [3] proposed a second alternative called the "u" test which is of the form

$$u = \frac{|\overline{X} - \mu|}{\dfrac{R}{d_2\sqrt{n}}},$$

where d_2 is the ratio of the expected value of the range to the population standard deviation which was tabulated by Tippett [4] and can be found tabulated in any quality control handbook. This form could also be used if the sample were divided into random sub-groups of equal size. The statistic is then

$$u = \frac{|\overline{X} - \mu|}{\dfrac{\overline{R}}{d_2\sqrt{nm}}},$$

where $\overline{R}$ is the average range of m sub-groups of n each.

Lord [3] also proposed a two-sample test of the form

$$u = \frac{|\overline{X}_1 - \overline{X}_2| \, d_2}{\overline{R}\sqrt{\dfrac{1}{nm_1} + \dfrac{1}{nm_2}}},$$

where the sample from which $\overline{X}_1$ was obtained is made up of m_1 sub-groups of n each and the sample from which $\overline{X}_2$ was obtained is made of m_2 subgroups, also of n each. $\overline{R}$ is the average range of the combined samples. The power of these tests has been discussed by Lord [2]

Use of Tables: All u-tests may be converted into simpler G-tests by suitable transformations The use of such G-tests effects a considerable saving in computing time, particularly for the two-sample test.

The significance of the difference between a sample mean $\overline{X}$ and a hypothetical value μ may be determined using a G-test in the form

$$G_1 = \frac{|\overline{X} - \mu|}{\overline{R}} = \frac{u}{d_2\sqrt{nm}}$$

Table I contains percentage points of the distribution of G_1, for significance levels $\alpha = .10$, $.05$, and $.01$; $n = 2, 3, \cdots, 15$, $m = 1, 2, \cdots, 15$. The special case $m = 1$, with a single range R in a sample of n values, is the original G-test of Daly.

The G-test for assessing the significance of the difference between two sample means takes the form

$$G_2 = \frac{|\overline{X}_1 - \overline{X}_2|}{\overline{R}} = \frac{u\sqrt{\dfrac{1}{nm_1} + \dfrac{1}{nm_2}}}{d_2}$$

Table II, in a quadruple entry form, gives the percentage points of the

TABLE I

PERCENTAGE POINTS FOR "G" TEST WHEN COMPARING A SAMPLE MEAN TO A HYPOTHETICAL MEAN

$$G_1 = \frac{|\overline{X} - \mu|}{\overline{R}}$$

m = number of subgroups
n = subgroup size

n	α \ m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	.10	3.16	1.16	.80	.64	.54	.48	.44	.40	38	35	33	.32	.30	.29	.28
2	.05	6.36	1.72	1.08	.83	.70	.61	.55	50	46	44	41	.39	.37	.36	.34
	01	31.84	3.96	1 99	1.39	1.10	.93	.82	74	67	.62	.58	55	.52	.50	.47
	10	.88	.49	37	.31	.27	.24	22	.21	.19	.18	17	17	.16	15	15
3	05	1.30	.64	.47	.38	.33	30	27	.25	24	.22	21	20	19	18	.18
	.01	3.01	1.05	.71	.56	.48	.42	38	35	33	.31	29	.27	26	.25	.24
	.10	.53	.32	.25	.21	.19	17	16	.15	14	.13	.12	.12	.11	.11	.10
4	.05	.72	.41	.31	.26	.23	.21	.19	.18	.17	.16	15	.14	.14	.13	.13
	.01	1.32	.62	.45	37	.32	.28	.26	24	22	.21	20	.19	.18	.18	.17
	.10	.39	.25	19	.17	.15	.13	.12	.11	11	.10	10	.09	.09	.09	08
5	.05	.51	.31	.24	.20	.18	.16	15	.14	.13	.12	12	11	.11	.10	10
	.01	.84	45	.34	.28	24	.22	.20	.19	.17	.16	.16	.15	.14	.14	.13
	.10	31	20	.16	.14	.12	.11	10	.10	.09	.08	.08	.08	.07	.07	.07
6	05	.40	.25	20	.17	.15	13	12	.11	11	.10	.10	09	.09	.09	.08
	.01	.63	.36	27	23	.20	18	17	.15	14	.14	13	.12	12	.11	11
	.10	.26	17	14	12	.11	10	09	.08	08	.07	.07	.07	.06	.06	.06
7	.05	33	21	.17	.14	.13	12	.11	10	.09	.09	.08	.08	.08	.07	.07
	.01	51	.30	.23	.19	.17	16	.14	13	12	.12	.11	.11	.10	10	.09
	.10	.23	15	.12	10	09	09	.08	.07	07	.07	06	06	.06	.06	.05
8	.05	29	.19	.15	.13	.11	.10	09	09	.08	.08	07	07	.07	.07	.06
	.01	.43	.26	.20	17	.15	.14	13	.12	11	.10	10	09	.09	09	.08
	.10	.20	..4	11	.09	.08	08	.07	.07	06	.06	06	.05	.05	.05	05
9	.05	.25	.17	.13	.11	.10	09	08	.08	.07	.07	.07	.06	.06	.06	.06
	.01	.37	.23	.18	.15	.14	.12	.11	.11	10	.09	09	.09	.08	.08	.08
	10	.19	12	.10	.09	.08	07	06	.06	06	05	.05	05	05	.05	.04
10	05	23	.15	12	10	.09	.08	.08	.07	.07	06	.06	.06	.06	.05	.05
	.01	.33	21	.16	.14	12	11	10	.10	.09	09	.08	.08	.07	.07	.07
	.10	.17	.12	.09	.08	.07	06	06	.06	.05	.05	.05	.05	.04	.04	.04
11	.05	.21	14	11	.10	.08	.08	07	.07	06	06	.06	.05	.05	.05	.05
	.01	.30	.19	.15	.13	11	.10	10	09	.08	.08	.08	.07	.07	.07	.06
	.10	16	11	09	.07	07	.06	06	.05	.05	05	.04	.04	.04	.04	.04
12	.05	.19	.13	10	.09	.08	.07	07	06	.06	.06	05	.05	.05	.05	.04
	.01	.28	.18	.14	12	11	.10	.09	.08	.08	07	.07	.07	.06	.06	.06
	.10	.15	.10	08	.07	06	.06	.05	.05	05	.04	04	.04	04	.04	.04
13	.05	.18	.12	.10	.08	.07	.07	.06	06	05	.05	05	.05	.05	.04	.04
	.01	26	17	.13	.11	.10	.09	.08	08	.07	.07	.07	.06	.06	.06	.06
	.10	.14	.09	.08	07	06	.05	.05	.05	.04	.04	.04	.04	.04	.03	03
14	.05	.17	.11	.09	.08	.07	.06	.06	.05	.05	.05	.05	.04	.04	.04	04
	.01	.24	.16	.12	.11	.09	08	.08	.07	.07	.06	.06	.06	06	.05	.05
	.10	.13	.09	.07	06	06	.05	.05	.04	.04	.04	04	.04	03	.03	.03
15	.05	.16	.11	09	.07	07	06	06	05	.05	05	04	04	04	.04	04
	.01	.22	.15	.12	.10	.09	.08	.07	07	.06	.06	06	.06	.05	.05	.05

TABLE II

PERCENTAGE POINTS FOR "G" TEST WHEN COMPARING INDEPENDENT SAMPLE MEANS

$$G_2 = \frac{|\overline{X}_1 - \overline{X}_2|}{\overline{R}}$$

m_1 = number of subgroups in sample from which $\overline{X}_1$ was obtained
m_2 = number of subgroups in sample from which $\overline{X}_2$ was obtained
n = subgroup size
These tables are symmetric.

Subgroup size n = 2

α	m_1 = 1	2	3	4	5	6	7	8	m_2
.10	2.32	1 69	1.47	1.36	1.30	1.25	1.22	1.20	1
.05	3.43	2 29	1 93	1.75	1.63	1.56	1.51	1.48	
.01	7.92	4.21	3.21	2.75	2.50	2.34	2.22	2.14	
		1.27	1.11	1.03	.97	.93	.90	.88	2
		1.67	1.42	1.29	1.21	1.16	1.12	1.09	
		2.78	2.25	1.98	1.81	1.70	1.62	1.56	
			.97	.89	.83	.80	.77	.75	3
			1.22	1.11	1.03	.98	.95	.92	
			1.86	1.65	1.52	1.43	1.36	1.31	
				.81	.76	.72	.70	.68	4
				1.00	.93	.89	.85	.83	
				1.47	1.35	1.27	1.21	1.17	
					.71	.67	.65	.63	5
					.87	.82	.79	.76	
					1.24	1.17	1.12	1.07	
						64	.61	.59	6
						.78	.74	.72	
						1.10	1.04	1 00	
							.58	.56	7
							.71	.68	
							.99	.95	
								.54	8
								.66	
								.91	

m_2	α								m_2
15	10	.39							
	.05	.46							
	.01	.63							
14	.10	.39	.40						
	.05	.47	.48						
	.01	.64	.65						
13	.10	.40	.41	.42					
	.05	.48	.49	.50					
	.01	.66	.67	.68					
12	.10	.41	.42	.43	.43				
	.05	.50	.50	.51	.52				
	.01	.67	.68	.70	.71				
11	.10	.42	.43	.44	.44	.45			
	,05	.51	.52	.53	.54	.55			
	01	.69	.70	.72	.73	.75			
10	.10	43	.44	.45	.46	.47	.48		
	.05	52	.53	.54	.55	.56	.58		
	.01	.71	.72	.74	.75	.77	.80		
9	.10	.45	.46	.46	.47	.48	.49	.51	
	.05	.54	.55	.56	.57	.58	.60	.61	
	.01	.74	.75	.76	78	.80	.82	.85	
8	.10	.47	.47	.48	.49	50	.51	.52	8
	.05	56	.57	.58	.59	.60	.62	.64	
	.01	.77	.78	.79	.81	.83	.85	.88	
7	.10	.49	.50	.50	.51	.52	.53	.55	7
	.05	.59	.60	.61	.62	.63	.65	.66	
	.01	.81	.82	.83	85	.87	.89	.92	
6	.10	.52	.52	.53	.54	.55	.56	.57	6
	.05	.62	63	.64	.65	.66	.68	70	
	.01	.85	.87	.88	.90	.92	.94	.97	
5	.10	.55	.56	.57	.57	.58	.59	.61	5
	.05	.67	.68	.69	.70	.71	.72	.74	
	.01	.92	.93	.95	.96	.98	1.01	1.04	
4	.10	.60	.61	.62	.63	.63	.65	.66	4
	.05	.73	.74	.75	.76	.77	.79	.80	
	.01	1.01	1.02	1.04	1.05	1.07	1.10	1.13	
3	.10	.68	.69	.69	.70	.71	.72	.74	3
	.05	.82	.83	.84	.85	.87	.88	.90	
	.01	1.14	1.15	1.17	1.19	1.21	1.23	1.27	
2	.10	.81	.82	.82	.83	.84	.85	.87	2
	.05	.99	.99	1.00	1.02	1.03	1.04	1.06	
	.01	1.36	1.38	1.39	1.42	1.44	.148	1.51	
1	.10	1.12	1.12	1.13	1.14	1.15	1.16	1.18	1
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	

TABLE II—*Continued*

Subgroup size n = 3

m_2	α	m_1 = 1	2	3	4	5	6	7	8
1	.10	.97	.79	.71	.67	.65	.64	.63	.62
	.05	1.27	.99	.89	.84	.80	.78	.76	.75
	.01	2.09	1.50	1.30	1.20	1.13	1.09	1.06	1.04
2	.10		.62	.55	.52	.49	.48	.47	.46
	.05		.77	.68	.63	.60	.58	.57	.56
	.01		1.13	.98	.90	.84	.81	.78	.77
3	.10			.49	.45	.43	.41	.40	.39
	.05			.60	.55	52	.50	.49	.47
	.01			.84	.77	.72	.69	.67	.65
4	.10				.41	.39	.38	.36	.35
	.05				.50	.48	.46	.44	.43
	.01				.70	.66	.63	.60	.58
5	.10					.37	.35	.34	.33
	.05					.45	.42	.41	.40
	.01					.61	.58	.56	.54
6	.10						.33	.32	.31
	.05						.40	.39	.37
	.01						.55	.53	.51
7	.10							.31	.30
	.05							.37	.36
	.01							.50	.49
8	.10								.29
	.05								.34
	.01								.47

m_2	α	m_1 = 15	14	13	12	11	10	9
15	.10	.21						
	.05	.25						
	.01	.33						
14	.10	.21	.21					
	.05	.25	.26					
	.01	.34	.34					
13	.10	.21	.22	.22				
	.05	.26	.26	.27				
	.01	.35	.35	.36				
12	.10	.22	.22	.23	.23			
	.05	.26	.27	.27	.28			
	.01	.35	.36	.37	.37			
11	.10	.22	.23	.23	.24	.24		
	.05	.27	.27	.28	.28	.29		
	01	.36	.37	.37	.38	.39		
10	.10	.23	.24	.24	.24	.25	.25	
	.05	.28	.28	.29	.29	.30	.31	
	.01	.37	.38	.39	.39	.40	.41	
9	.10	.24	.24	.25	.25	.26	.26	.27
	.05	.29	.29	.30	.30	31	.31	.32
	.01	.39	.39	.40	.41	.41	.42	.44
8	.10	.25	.25	.26	.26	.26	.27	.28
	.05	30	.30	31	.31	32	.33	.33
	.01	.40	.41	.41	.42	.43	.44	.45
7	.10	.26	.26	.27	.27	.28	.28	.29
	.05	.31	.32	.32	.33	.33	.34	.35
	.01	.42	.43	.43	.44	.45	.46	.47
6	.10	.28	.28	.28	.28	.29	.30	.30
	.05	.33	.33	.34	.34	.35	.36	.36
	.01	.44	.45	.46	.46	.47	.48	.49
5	.10	.29	.30	.30	.30	.31	.32	.32
	.05	.35	.36	.36	.37	.37	.38	.39
	.01	.48	.48	.49	.50	.50	.51	.52
4	.10	.32	.32	.33	.33	.34	.34	.35
	.05	.39	.39	.39	.40	.40	.41	.42
	.01	.52	.53	.53	.54	.55	.56	.57
3	.10	.36	.36	.37	.37	.38	.38	.38
	.05	.43	.44	.44	.45	.45	.46	.46
	.01	.59	.59	.60	.61	.61	.62	.63
2	.10	.43	.43	.44	.44	.44	.45	.45
	.05	.52	.52	.53	.53	.53	.54	.55
	.01	.70	.71	.71	.72	.73	.74	.75
1	.10	.59	.60	.60	.60	.60	.61	.61
	.05	.71	.72	.72	.72	.73	.73	.74
	.01	.97	.97	.98	.99	.99	1.01	1.02

TABLE II—*Continued*

Subgroup size n = 4

m_2	α	m_1 = 1	2	3	4	5	6	7	8	m_1 = 15	14	13	12	11	10	9	m_2
	.10	.65	.53	.49	.47	.46	.45	.44	.44								
	.05	.81	.66	.60	.57	.55	.54	.53	.53								1
	.01	1.24	.95	85	.79	.76	.74	.72	.71								
	.10		.42	.38	.36	.35	.34	.33	.32								
	.05		.52	.47	.44	.42	.41	.40	.39								2
	.01		.73	.65	.60	.57	.55	.54	.53								
	.10			.34	.32	.30	.29	.28	.28	.15							
15	.05			.41	.38	.36	.35	.34	.33	.18							3
	.01			.57	.52	.49	.48	.46	.45	.23							
	.10				.29	.28	.26	.26	.25	.15	.15						
14	.05				.35	.33	.32	.31	.30	.18	.18						4
	.01				.48	.45	.43	.42	.40	.24	.24						
	10					.26	.25	.24	23	.15	.16	.16					
13	.05					.31	.30	.29	.28	.18	.19	.19					5
	.01					.42	.40	.39	.38	.24	.25	.25					
	.10						.24	.23	.22	.16	.16	.16	16				
12	.05						.28	.27	.26	.19	19	.19	.20				6
	.01						.38	.37	.35	.25	25	.26	.26				
	10							.22	.21	.16	.16	.17	.17	.17			
11	05							.26	.25	19	19	20	.20	.21			7
	.01							.35	.34	.25	.26	.26	.27	.28			
	.10								.20	.16	.17	.17	.17	.18	.18		
10	.05								.24	.20	.20	.20	.21	.21	.22		8
	01								.33	26	.27	27	.28	.28	.29		
	.10									.17	.17	.17	18	.18	.19	.19	
9	.05									.20	.21	.21	.21	.22	.22	.23	
	.01									.27	.27	.28	28	.29	.30	.31	m_2
	.10									.18	.18	.18	.19	.19	19	.20	
8	.05									.21	.21	.22	.22	.23	23	.24	8
	.01									.28	.29	.29	.29	.30	31	32	
	.10									.18	.19	.19	.19	.20	.20	.20	
7	.05									.22	.22	.23	.23	.23	24	.24	7
	.01									.30	.30	.30	.31	.31	.32	.33	
	10									.20	.20	.20	.20	.21	.21	.21	
6	.05									.23	.24	.24	.24	.25	.25	.26	6
	.01									.31	.32	.32	.32	.33	.34	.34	
	.10									.21	.21	.21	.22	.22	22	.23	
5	.05									.25	.25	.26	.26	.26	.27	.27	5
	.01									.33	.34	.34	.35	.35	.36	37	
	.10									.23	.23	.23	.23	24	.24	25	
4	.05									27	.28	.28	28	.29	.29	.29	4
	01									36	.37	.37	.38	.38	.39	.40	
	.10									.26	26	.26	.26	.27	.27	.27	
3	.05									.31	.31	31	.32	.32	.32	.33	3
	.01									.41	.41	.42	.42	.43	.43	.44	
	.10									.31	.31	.31	.31	.31	.32	.32	
2	.05									.37	37	.37	.37	.38	.38	.38	2
	.01									.49	49	.49	.50	.51	.51	.52	
	.10									.42	.42	.42	.42	43	.43	.43	
1	.05									50	.51	.51	.51	.51	51	.52	1
	.01									.67	.67	.68	.69	.69	.70	.70	
m_2	α									15	14	13	12	11	10	9	m_2

m_1

TABLE II—*Continued*

Subgroup size n = 5

	α	1	2	3	4	5	6	7	8	m_1 / m_2
	.10	.49	.41	.38	.37	.36	.35	.35	.34	
	.05	.61	.51	.47	.45	.43	.42	.42	.41	1
	.01	.89	.71	.64	.61	.59	.57	.56	.55	
	.10		.33	.30	.28	.27	.26	.26	.26	
	.05		.40	.36	34	.33	.32	.31	.31	2
	.01		.56	.50	.46	.44	.43	.42	.41	
	.10	.12		.27	.25	.24	.23	.22	.22	
15	.05	.14		.32	.30	.29	.27	.27	.26	3
	.01	.18		.44	.40	38	.37	.36	.35	
	.10	.12	.12		.23	.22	.21	.20	.20	
14	.05	.14	.14		.28	.26	.25	.24	.24	4
	.01	.19	.19		.37	.35	.34	.32	.32	
	.10	.12	.12	.12		.20	.19	.19	.18	
13	.05	.14	.15	.15		.24	.23	.23	.22	5
	.01	.19	.19	.20		.33	.31	.30	.29	
	.10	.12	.12	.13	.13		.18	.18	.17	
12	.05	.15	.15	.15	.16		.22	.21	.21	6
	.01	.20	.20	.20	.21		.30	.29	.28	
	.10	.13	.13	.13	.13	.14		.17	.17	
11	.05	.15	.15	.16	.16	.16		.21	.20	7
	.01	.20	.20	.21	.21	.22		.27	.26	
	.10	.13	.13	.13	.14	14	.14		.16	
10	.05	.16	.16	.16	.16	.17	.17		.19	8
	.01	.21	.21	.21	.22	.22	.23		.25	
	.10	.13	.14	.14	.14	14	.15	.15		
9	.05	.16	.16	.17	.17	.17	.18	.18		
	.01	.21	.22	.22	.22	.23	.23	.24		
	.10	.14	.14	.14	.15	.15	.15	.16		
8	.05	.17	.17	.17	.17	.18	.18	.19		8
	.01	.22	.22	.23	.23	.24	.24	.25		
	.10	.15	15	15	.15	.15	16	.16		
7	.05	.18	.18	.18	.18	.19	.19	.19		7
	.01	23	.23	.24	.24	.25	25	.26		
	.10	.15	.16	.16	.16	.16	.16	.17		
6	.05	.18	.19	.19	.19	.19	.20	.20		6
	.01	.24	.25	.25	.25	.26	.26	.27		
	.10	.16	.17	.17	17	.17	.18	.18		
5	.05	.20	.20	.20	.20	.21	.21	.21		5
	.01	.26	.26	.27	.27	.27	.28	.29		
	.10	.18	.18	.18	.18	.19	.19	.19		
4	.05	.22	.22	.22	.22	.22	.23	.23		4
	.01	.29	.29	.29	.29	.30	.30	.31		
	.10	.20	.20	.20	.21	.21	.21	.21		
3	.05	.24	.24	.25	.25	.25	.25	.26		3
	.01	.32	.32	.33	.33	.33	.34	.34		
	.10	.24	.24	.24	.24	.25	.25	.25		
2	.05	.29	.29	.29	.29	.30	.30	.30		2
	.01	.38	.39	.39	.39	.39	.40	.40		
	.10	.33	.33	.33	.33	.33	.34	.34		
1	.05	.40	.40	.40	.40	.40	.41	.41		1
	.01	.53	53	53	.53	.54	.54	.55		
m_2	α	15	14	13	12	11	10	9		m_2

m_1

TABLE II—*Continued*

Subgroup size n = 6

m_2	α	m_1 = 1	2	3	4	5	6	7	8
1	.10	.41	.34	.32	.31	.30	.29	.29	.29
	.05	.50	.42	.39	.37	.36	.35	.35	.34
	.01	.71	.58	.53	.50	.48	.47	.46	.46
2	.10		.28	.25	.24	.23	.22	.22	.21
	.05		.33	.30	.28	.27	.26	.26	.25
	.01		.46	.41	.38	.37	.36	.35	.34
3	.10			.22	.21	.20	.19	.19	.18
	.05			.27	.25	.24	23	,22	.22
	.01			.36	.33	.32	.31	.30	.29
4	.10				.19	.18	.17	.17	.16
	.05				.23	.22	.21	.20	.20
	.01				.31	.29	.28	.27	.26
5	.10					.17	.16	.16	.15
	.05					.20	.20	.19	.18
	.01					.27	.26	.25	.24
6	.10						.15	.15	.14
	.05						.19	.18	.17
	.01						.25	.24	.23
7	.10							.14	.14
	.05							.17	.17
	.01							.23	.22
8	.10								.13
	.05								.16
	.01								.21

m_2	α	m_1 = 15	14	13	12	11	10	9	m_2
15	.10	.10							
	.05	.12							
	.01	.15							
14	.10	.10	10						
	.05	.12	.12						
	.01	.16	.16						
13	.10	.10	.10	.10					
	.05	.12	.12	.12					
	.01	.16	.16	.17					
12	.10	.10	.10	.11	.11				
	.05	.12	.13	.13	.13				
	.01	.16	.17	.17	.17				
11	.10	.11	.11	.11	.11	.11			
	.05	.13	.13	.13	.13	.14			
	.01	.17	.17	.17	18	.18			
10	.10	.11	.11	.11	.11	.12	.12		
	.05	.13	.13	.13	.14	.14	.14		
	.01	.17	.17	.18	.18	.18	.19		
9	.10	.11	.11	.12	.12	.12	.12	.13	
	.05	.13	14	14	.14	.14	.15	.15	
	.01	.18	18	.18	.19	19	.19	.20	
8	.10	.12	12	.12	.12	.12	13	.13	8
	.05	.14	.14	.14	.15	.15	.15	16	
	.01	.18	.19	.19	.19	.20	.20	.21	
7	.10	.12	.12	.13	.13	.13	.13	13	7
	.05	.15	.15	.15	.15	.15	.16	.16	
	.01	.19	.20	.20	.20	.20	.21	.21	
6	.10	.13	.13	.13	.13	.14	.14	.14	6
	.05	.15	.16	.16	.16	16	.16	.17	
	.01	.20	.21	.21	.21	.22	.22	.22	
5	.10	.14	.14	.14	.14	.14	.15	.15	5
	.05	.16	.17	.17	.17	.17	.18	.18	
	.01	.22	.22	.22	.23	.23	.23	.24	
4	.10	.15	.15	.15	.15	.16	.16	.16	4
	.05	.18	.18	.18	.18	.19	.19	.19	
	.01	.24	.24	.24	.25	.25	.25	.26	
3	.10	.17	.17	.17	.17	.17	.18	.18	3
	.05	.20	.20	.20	.21	.21	.21	.21	
	.01	.27	.27	.27	.27	.28	.28	.29	
2	.10	.20	.20	.20	.20	.21	.21	.21	2
	.05	.24	.24	.24	.24	.25	.25	.25	
	.01	.32	.32	.32	.33	.33	.33	.34	
1	.10	.28	.28	.28	.28	.28	.28	.28	1
	.05	.33	.33	.33	.33	.34	.34	.34	
	.01	.44	44	.44	.44	.45	.45	.45	

TABLE II—*Continued*

Subgroup size n = 7

m_2	α	m_1 = 1	2	3	4	5	6	7	8
1	.10	.35	.29	27	26	.26	.25	.25	.25
	.05	.43	.36	.33	.32	.31	.31	30	.30
	01	.60	.49	.45	43	.42	.41	.40	.40
2	.10		.24	.22	20	.20	.19	19	.18
	.05		.29	26	.24	.24	.23	23	.22
	.01		.39	.35	.33	.32	.31	.30	.29
3	.10			.19	.18	.17	.16	.16	.16
	.05			.23	.22	.21	.20	.19	.19
	.01			31	.29	27	.26	.26	.25
4	.10				16	.16	.15	.15	.14
	.05				.20	.19	.18	.18	.17
	.01				.26	.25	.24	.23	.23
5	10					.15	.14	.14	.13
	05					.18	.17	.16	.16
	.01					.24	.22	.22	.21
6	.10						.13	.13	13
	05						.16	.15	.15
	.01						.21	.21	.20
7	.10							.12	.12
	.05							.15	.14
	.01							.20	.19
8	,10								.12
	05								.14
	.01								.18

m_2	α	m_1 = 15	14	13	12	11	10	9
15	.10	.08						
	.05	.10						
	.01	13						
14	.10	.09	.09					
	.05	.10	.10					
	.01	.14	.14					
13	10	.09	.09	.09				
	05	.10	.11	11				
	.01	14	14	.14				
12	.10	09	.09	.09	.09			
	05	.11	.11	.11	.11			
	.01	.14	.14	.15	15			
11	.10	09	.09	.09	10	.10		
	.05	11	11	.11	12	.12		
	.01	.14	15	.15	15	.16		
10	,10	.09	.10	.10	.10	10	.10	
	05	.11	.11	12	.12	.12	.12	
	.01	.15	.15	.15	.16	.16	.16	
9	.10	.10	.10	10	10	.10	.11	.11
	.05	12	.12	.12	.12	.12	.13	.13
	.01	.15	.16	.16	.16	16	.17	17
8	.10	.10	10	.10	.11	11	.11	.11
	.05	12	12	12	.13	.13	.13	.13
	01	16	.16	.16	17	17	17	.18
7	.10	11	11	.11	11	11	11	.12
	.05	.13	.13	.13	13	13	.14	14
	01	.17	.17	.17	17	18	.18	.19
6	10	11	11	11	12	12	12	12
	.05	13	14	14	14	.14	14	15
	.01	18	18	.18	.18	.19	.19	.19
5	10	.12	12	12	12	13	13	.13
	.05	14	.14	.15	15	.15	15	.16
	.01	.19	19	.19	.20	20	20	21
4	.10	13	13	.13	13	14	.14	14
	.05	.16	.16	16	.16	.16	.16	.17
	.01	21	.21	.21	.21	21	.22	22
3	.10	.15	.15	.15	.15	15	.15	15
	05	.18	.18	18	.18	.18	18	.19
	01	23	.23	24	.24	24	.24	.25
2	.10	.17	18	.18	18	.18	.18	.18
	.05	.21	.21	21	21	21	.22	.22
	.01	.28	.28	.28	.28	.28	.29	.29
1	.10	.24	.24	.24	.24	.24	.24	.25
	.05	29	.29	.29	.29	.29	.29	.29
	.01	.38	.38	.38	.38	.39	.39	.39

TABLE II—*Continued*

Subgroup size n = 6

m_2	α	m_1 = 15	14	13	12	11	10	9	m_1 = 1	2	3	4	5	6	7	8	m_2
	.10								.31	.26	.24	.23	.23	.22	.22	.22	1
	.05								.37	.31	.29	.28	.27	.27	.27	.26	
	.01								.52	.43	.40	.38	.37	.36	.35	.35	
	.10									.21	.19	.18	.17	.17	.17	.16	2
	.05									.25	.23	.22	.21	.20	.20	.20	
	01									.34	.31	.29	.28	.27	.26	.26	
15	.10	.07									.17	.16	.15	.15	.14	.14	3
	.05	.09									.20	.19	.18	.18	.17	17	
	.01	.12									.27	.25	.24	.23	.23	.22	
14	.10	.08	.08									.15	.14	.13	.13	.13	4
	.05	.09	.09									.18	.17	16	.15	.15	
	.01	.12	.12									.23	.22	.21	.21	.20	
13	.10	.08	.08	.08									.13	.12	12	.12	5
	.05	.09	.09	.10									.16	.15	.14	.14	
	.01	.12	.12	13									.21	20	.19	.19	
12	.10	.08	.08	.08	.08									.12	.11	.11	6
	.05	.09	.10	.10	10									.14	.14	.13	
	.01	.13	.13	.13	.13									.19	.18	.18	
11	10	.08	.08	.08	.09	.09									.11	.11	7
	.05	.10	.10	.10	.10	.10									.13	.13	
	.01	.13	.13	.13	.14	.14									.17	.17	
10	.10	.08	.08	.09	.09	.09	.09									.10	8
	.05	.10	.10	.10	.11	.11	.11									.12	
	.01	.13	.13	.14	.14	.14	.14									.16	
9	.10	.09	.09	.09	.09	.09	.09	.10									
	.05	.10	.10	.11	11	.11	.11	.12									
	.01	.14	.14	.14	.14	.15	.15	.15									
8	.10	.09	.09	.09	.09	.10	.10	.10									8
	05	.11	.11	11	.11	.11	12	.12									
	.01	.14	.14	.15	15	.15	.15	16									
7	.10	.09	09	.10	.10	.10	10	.10									7
	.05	.11	.11	.12	12	.12	.12	12									
	01	.15	.15	.15	.15	.16	.16	.16									
6	.10	.10	10	.10	.10	.10	.11	.11									6
	.05	.12	.12	.12	12	.12	.13	.13									
	01	.16	.16	.16	.16	.17	.17	.17									
5	.10	11	.11	.11	.11	.11	.11	.12									5
	.05	.13	.13	13	.13	.13	.13	.14									
	.01	.17	.17	17	17	.18	.18	.18									
4	.10	.12	.12	.12	.12	.12	.12	.12									4
	.05	.14	14	.14	.14	.14	.15	15									
	.01	.18	18	.19	.19	.19	.19	.20									
3	.10	.13	.13	.13	.13	.13	14	.14									3
	.05	.16	.16	.16	.16	.16	.16	.16									
	.01	.21	.21	.21	.21	.21	.22	.22									
2	.10	.16	.16	.16	.16	.16	.16	.16									2
	.05	.19	.19	.19	.19	.19	.19	.19									
	.01	.24	.25	.25	.25	.25	.25	.26									
1	.10	.21	.21	.21	.21	.22	.22	.22									1
	.05	.25	.25	.26	.26	.26	.26	.26									
	.01	.34	.34	.34	.34	34	.35	.35									

TABLE II—*Continued*

Subgroup Size $n = 9$

m_2	α	m_1 = 1	2	3	4	5	6	7	8
1	.10	.27	.23	22	.21	.21	.20	.20	.20
	.05	.33	.28	.26	.25	.25	.24	.24	.24
	.01	.46	38	.36	.34	33	.32	.32	.32
2	.10		.19	.17	.16	.16	.15	15	.15
	.05		.23	.21	.20	.19	.18	18	.18
	.01		.31	.28	26	.25	.24	.24	.23
3	.10			.15	.14	14	.13	.13	.13
	.05			.18	.17	.16	.16	.15	.15
	.01			.25	.23	.22	.21	.20	.20
4	.10				.13	.13	.12	.12	.11
	.05				.16	.15	14	.14	.14
	.01				.21	.20	.19	.19	.18
5	.10					.12	11	.11	.11
	.05					.14	.14	.13	.13
	.01					.19	.18	17	.17
6	.10						.11	.10	.10
	.05						.13	.12	.12
	.01						.17	.16	.16
7	.10							.10	10
	.05							.12	12
	.01							.16	.15
8	.10								.09
	.05								.11
	.01								.15

m_2	α	m_1 = 15	14	13	12	11	10	9
15	.10	.07						
	.05	.08						
	.01	.11						
14	.10	.07	.07					
	.05	.08	.08					
	.01	.11	.11					
13	.10	.07	.07	.07				
	.05	.08	09	.09				
	.01	.11	.11	.11				
12	.10	.07	.07	.07	.08			
	.05	.09	.09	.09	.09			
	.01	.11	.12	.12	.12			
11	10	.07	07	.08	.08	.08		
	.05	09	.09	.09	.09	.09		
	.01	.12	.12	.12	.12	.12		
10	.10	.08	.08	.08	.08	.08	.08	
	.05	.09	.09	.09	.10	.10	.10	
	.01	.12	.12	.12	.13	.13	.13	
9	.10	.08	.08	.08	.08	.08	09	.09
	.05	.09	.10	.10	.10	.10	.10	.10
	.01	.12	13	.13	.13	.13	.13	.14
8	.10	.08	.08	.08	08	.09	.09	.09
	.05	.10	10	.10	.10	.10	.11	.11
	01	.13	.13	.13	13	.14	.14	.14
7	.10	.08	.09	.09	.09	.09	.09	.09
	.05	.10	.10	.10	.11	.11	.11	.11
	.01	.13	.14	.14	.14	.14	.14	.15
6	.10	.09	09	.09	.09	.09	.10	.10
	05	.11	.11	.11	.11	.11	11	.12
	.01	.14	14	14	.15	.15	15	.16
5	.10	.10	.10	.10	.10	.10	.10	.10
	.05	.11	.12	.12	.12	.12	.12	.12
	01	.15	.15	.15	.16	.16	.16	.16
4	.10	.10	.11	.11	.11	.11	.11	.11
	.05	.13	.13	.13	.13	.13	.13	.13
	.01	.16	.17	.17	.17	.17	.17	.18
3	.10	.12	.12	.12	.12	.12	.12	.12
	05	.14	.14	.14	.14	.14	.15	.15
	01	.19	.19	.19	.19	.19	19	.20
2	10	.14	.14	.14	14	.14	.14	.15
	.05	.17	.17	.17	17	.17	.17	.17
	.01	.22	.22	.22	.23	.23	.23	.23
1	.10	19	.19	.19	.19	.19	.20	.20
	.05	23	.23	123	.23	.23	.23	.24
	.01	.30	.30	31	.31	31	.31	.31

TABLE II—*Continued*

Subgroup size n = 10

m_2	α	1	2	3	4	5	6	7	8	m_2
	.10	.25	.21	.20	.19	.19	.18	.18	.18	
	.05	.30	.26	.24	.23	.22	.22	.22	.22	1
	.01	.42	.35	.32	.31	.30	.30	.29	.29	
	.10		.17	.16	.15	.14	.14	.14	.13	
	.05		.21	.19	18	.17	.17	.16	.16	2
	.01		.28	.25	.24	.23	.22	.22	.21	
	.10		.06	14	.13	12	.12	.12	.11	
15	.05		.07	.17	.16	15	.14	.14	.14	3
	.01		.10	22	.21	.20	.19	.19	.18	
	.10		06	.06	.12	.11	11	.11	.10	
14	.05		08	.08	.14	.14	.13	.13	.12	4
	.01		.10	.10	.19	18	.17	.17	.17	
	10		.06	.07	.07	.11	.10	.10	.10	
13	.05		.08	.08	.08	.13	.12	.12	.12	5
	.01		.10	10	.10	.17	.16	16	.15	
	.10		.07	.07	.07	.07	10	.09	.09	
12	.05		.08	.08	.08	.08	.12	.11	.11	6
	.01		.10	.11	.11	.11	.16	.15	.15	
	.10		.07	.07	.07	07	.07	.09	.09	
11	.05		.08	.08	.08	.08	.09	.11	.11	7
	.01		.11	.11	.11	.11	.11	.14	.14	
	.10		.07	.07	.07	.07	.07	.08	.08	
10	.05		.08	.08	.09	.09	.09	.09	.10	8
	.01		.11	.11	.11	11	.12	.12	.13	
	.10		.07	.07	.07	.07	.08	.08	.08	
9	.05		.09	.09	09	.09	.09	.09	.10	
	.01		.11	.11	.12	.12	.12	.12	.13	m_2
	.10		.07	.08	.08	.08	.08	.08	.08	
8	.05		.09	.09	.09	.09	.09	.10	.10	8
	.01		.12	.12	.12	.12	12	.13	.13	
	.10		08	.08	.08	.08	.08	.08	.09	
7	.05		.09	.09	.09	.10	.10	10	.10	7
	.01		.12	.13	.13	.13	.13	.13	.14	
	.10		.08	.08	.08	.08	.09	.09	.09	
6	.05		.10	.10	.10	.10	.10	.11	.11	6
	.01		.13	.13	.13	.13	.14	.14	.14	
	.10		09	09	.09	09	.09	.09	.09	
5	.05		.10	.11	.11	.11	11	.11	.11	5
	.01		.14	.14	.14	.14	.14	.15	.15	
	.10		.10	.10	.10	.10	.10	.10	.10	
4	.05		.11	.11	.12	.12	.12	.12	.12	4
	.01		.15	.15	.15	.15	.16	.16	.16	
	.10		.11	11	.11	.11	.11	.11	.11	
3	.05		.13	.13	.13	.13	.13	.13	.14	3
	01		.17	.17	.17	.17	.18	.18	.18	
	10		.13	.13	13	.13	.13	.13	.13	
2	.05		.15	.15	.15	.16	.16	.16	.16	2
	.01		.20	.20	.20	.21	.21	.21	.21	
	.10		.18	.18	.18	.18	.18	.18	.18	
1	.05		.21	.21	.21	.21	.21	.21	.21	1
	.01		.28	.28	.28	.28	.28	.28	.28	
	α		15	14	13	12	11	10	9	m_2

m_1

TABLE II—*Continued*

Subgroup size n = 11

m_2	α	m_1 = 1	2	3	4	5	6	7	8
1	.10	.23	.20	.18	.18	.17	.17	.17	.17
	.05	.28	.24	.22	.21	.21	.20	.20	.20
	.01	.38	.32	.30	.28	.28	.27	.27	.27
2	.10		.16	.14	.14	.13	.13	.13	.12
	.05		.19	.17	.16	.16	.15	.15	.15
	.01		.26	.23	.22	.21	.20	.20	.20
3	.10			.13	.12	.12	.11	.11	.11
	.05			.16	.14	.14	.13	.13	.13
	.01			.21	.19	.18	.18	.17	.17
4	.10				.11	.11	.10	.10	.10
	.05				.13	.13	.12	.12	.12
	.01				.18	.17	.16	.16	.15
5	.10					.10	.10	.09	.09
	.05					.12	.11	.11	.11
	.01					.16	.15	.15	.14
6	.10						.09	.09	.08
	.05						.11	.10	.10
	.01						.14	.14	.13
7	.10							.08	.08
	.05							.10	.10
	.01							.13	.13
8	.10								.08
	.05								.09
	.01								.12

m_2	α	m_1 = 15	14	13	12	11	10	9
15	.10	.06						
	.05	.07						
	.01	.09						
14	.10	.06	.06					
	.05	.07	.07					
	.01	.09	.09					
13	.10	.06	.06	.06				
	.05	.07	.07	.07				
	.01	.09	.10	.10				
12	.10	.06	.06	.06	.06			
	.05	.07	.07	.08	.08			
	.01	.10	.10	.10	.10			
11	.10	.06	.06	.06	.07	.07		
	.05	.07	.08	.08	.08	.08		
	.01	.10	.10	.10	.10	.11		
10	.10	.06	.07	.07	.07	.07	.07	
	.05	.08	.08	.08	.08	.08	.08	
	.01	.10	.10	.10	.11	.11	.11	
9	.10	.07	.07	.07	.07	.07	.07	.07
	.05	.08	.08	.08	.08	.08	.09	.09
	.01	.10	.11	.11	.11	.11	.11	.12
8	.10	.07	.07	.07	.07	.07	.07	.08
	.05	.08	.08	.08	.09	.09	.09	.09
	.01	.11	.11	.11	.11	.11	.12	.12
7	.10	.07	.07	.07	.07	.08	.08	.08
	.05	.09	.09	.09	.09	.09	.09	.09
	.01	.11	.11	.12	.12	.12	.12	.13
6	.10	.08	.08	.08	.08	.08	.08	.08
	.05	.09	.09	.09	.09	.09	.10	.10
	.01	.12	.12	.12	.12	.13	.13	.13
5	.10	.08	.08	.08	.08	.08	.09	.09
	.05	.10	.10	.10	.10	.10	.10	.10
	.01	.13	.13	.13	.13	.13	.14	.14
4	.10	.09	.09	.09	.09	.09	.09	.09
	.05	.11	.11	.11	.11	.11	.11	.11
	.01	.14	.14	.14	.14	.14	.15	.15
3	.10	.10	.10	.10	.10	.10	.10	.10
	.05	.12	.12	.12	.12	.12	.12	.13
	.01	.16	.16	.16	.16	.16	.16	.17
2	.10	.12	.12	.12	.12	.12	.12	.12
	.05	.14	.14	.14	.14	.14	.15	.15
	.01	.19	.19	.19	.19	.19	.19	.19
1	.10	.16	.16	.16	.16	.16	.17	.17
	.05	.19	.19	.20	.20	.20	.20	.20
	.01	.26	.26	.26	.26	.26	.26	.26

TABLE II—*Continued*

Subgroup Size n = 12

m_2	α	m_1 = 1	2	3	4	5	6	7	8
1	.10	.21	.18	.17	.17	.16	.16	.16	.16
	.05	.26	.22	21	.20	.19	.19	.19	.19
	.01	.35	30	.28	.27	.26	.25	.25	.25
2	.10		.15	.13	.13	.12	.12	.12	.12
	.05		.18	.16	.15	.15	.14	.14	.14
	.01		.24	.22	.20	20	.19	.19	.18
3	.10			.12	.11	.11	.10	.10	.10
	.05			.14	.13	.13	12	.12	.12
	.01			.19	.18	.17	.16	.16	16
4	.10				.10	.10	.09	.09	.09
	.05				.12	.12	.11	.11	.11
	.01				.17	.16	.15	.15	.14
5	.10					.09	.09	.09	.08
	.05					.11	.11	10	.10
	.01					15	.14	.14	.13
6	.10						.08	.08	.08
	.05						.10	.10	.09
	.01						.13	.13	.12
7	.10							.08	.08
	.05							.09	.09
	.01							.12	.12
8	.10								.07
	.05								.09
	.01								.12

m_2	α	m_1 = 15	14	13	12	11	10	9
15	.10	.05						
	.05	.06						
	.01	.08						
14	.10	.05	.06					
	.05	.06	.07					
	.01	.09	.09					
13	.10	.06	06	.06				
	.05	07	07	.07				
	.01	.09	.09	.09				
12	.10	.06	.06	.06	.06			
	.05	.07	.07	.07	.07			
	.01	.09	.09	.09	.09			
11	.10	.06	.06	.06	.06	.06		
	.05	.07	.07	.07	.07	.07		
	.01	.09	.09	.09	.10	.10		
10	.10	.06	.06	.06	.06	.06	07	
	.05	.07	.07	.07	.07	.08	.08	
	.01	.09	.10	.10	.10	.10	.10	
9	.10	.06	.06	.06	06	.07	.07	07
	05	.07	.07	.08	.08	.08	.08	.08
	.01	.10	.10	.10	.10	.10	.11	.11
8	.10	.06	.06	.07	.07	.07	.07	.07
	05	.08	.08	.08	.08	.08	08	.08
	.01	.10	10	.10	11	.11	.11	11
7	.10	.07	.07	.07	.07	.07	.07	.07
	05	.08	.08	.08	.08	.08	.c9	.09
	.01	.11	.11	.11	11	.11	.11	12
6	10	.07	07	.07	.07	.07	.08	.08
	,05	.08	.09	.09	.09	.09	.09	.09
	.01	.11	.11	.11	.12	.12	.12	.12
5	.10	.08	.08	.08	.08	.08	.08	.08
	.05	.09	.09	.09	.09	09	.10	.10
	.01	.12	.12	.12	.12	.12	.13	13
4	.10	.08	.08	.08	.08	.09	.09	.09
	05	.10	10	.10	.10	.10	.10	.11
	01	.13	.13	.13	.13	.13	.14	.14
3	.10	.09	.09	.09	.09	.10	.10	.10
	.05	.11	.11	.11	.11	.11	.12	12
	01	.15	15	.15	.15	.15	.15	.15
2	.10	.11	11	.11	.11	.11	.11	.11
	05	.13	13	.13	.13	.13	.14	14
	.01	.17	.17	.18	.18	.18	.18	.18
1	.10	.15	.15	.15	.15	.15	.15	.16
	.05	.18	.18	.18	.18	.18	.18	18
	01	.24	.24	.24	.24	.24	.24	24

TABLE II—*Continued*

Subgroup Size n = 13

m_2	α	m_1 = 1	2	3	4	5	6	7	8
1	10	.20	.17	16	.16	15	15	.15	15
	.05	.24	21	19	19	.18	.18	18	.17
	.01	33	28	26	25	24	.24	.23	.23
2	.10		14	13	12	.12	11	.11	11
	.05		.17	15	14	.14	.13	.13	13
	.01		.22	.20	.19	.18	.18	.17	17
3	.10			11	11	.10	.10	10	.09
	.05			.14	.13	.12	.12	.11	11
	.01			.18	.17	.16	15	15	.15
4	.10				10	.09	.09	09	.08
	.05				.12	.11	.11	.10	10
	.01				.15	15	14	.14	.13
5	.10					.09	08	08	.08
	.05					.10	.10	.10	.09
	.01					.14	.13	.13	12
6	.10						.08	08	.07
	.05						.09	.09	.09
	.01						.13	.12	.12
7	.10							.07	.07
	.05							.09	.08
	.01							12	11
8	.10								.07
	.05								.08
	.01								.11

m_2	α	m_1 = 15	14	13	12	11	10	9	m_2
15	.10	.05							
	.05	.06							
	.01	.08							
14	.10	05	.05						
	.05	.06	.06						
	.01	.08	.08						
13	.10	.05	05	.05					
	.05	.06	06	.06					
	.01	.08	.08	.08					
12	.10	.05	.05	.05	.06				
	.05	.06	.06	.07	.07				
	.01	.08	.08	.09	.09				
11	.10	.05	06	.06	.06	.06			
	.05	07	07	07	.07	.07			
	.01	09	09	.09	.09	.09			
10	.10	.06	.06	.06	06	06	.06		
	05	.07	.07	07	.07	07	.07		
	.01	.09	09	.09	.09	.09	.10		
9	.10	.06	06	.06	.06	.06	.06	.06	
	.05	.07	.07	.07	.07	.07	.08	.08	
	.01	.09	09	.09	.10	.10	.10	.10	
8	.10	.06	06	06	.06	06	.07	.07	8
	.05	07	07	.07	07	08	.08	.08	
	.01	.09	10	.10	.10	.10	.10	11	
7	.10	.06	.06	.06	07	07	.07	.07	7
	.05	.07	.08	.08	.08	.08	.08	.08	
	.01	.10	.10	10	.10	10	.11	.11	
6	.10	07	07	.07	.07	.07	.07	.07	6
	.05	.08	.08	08	08	.08	.08	.09	
	.01	10	11	.11	.11	11	.11	.11	
5	.10	.07	.07	.07	07	.07	08	.08	5
	.05	.08	.09	.09	.09	.09	.09	09	
	.01	.11	.11	.11	.11	.12	.12	.12	
4	.10	.08	.08	.08	.08	08	.08	.08	4
	.05	.09	.09	.09	.09	.10	.10	10	
	.01	.12	.12	.12	.12	.13	13	.13	
3	.10	.09	.09	.09	.09	.09	.09	.09	3
	.05	10	.10	.11	.11	.11	.11	.11	
	.01	.14	.14	.14	.14	.14	.14	.14	
2	.10	.10	.10	.10	.10	.11	.11	.11	2
	.05	.12	.12	12	.13	.13	.13	.13	
	.01	.16	.16	.16	.17	.17	.17	.17	
1	.10	.14	14	.14	.14	.14	.14	.15	1
	.05	17	.17	17	17	17	17	17	
	.01	.22	.22	.23	23	23	.23	23	

TABLE II—*Continued*

Subgroup Size n = 14

m_2	α	m_1 = 1	2	3	4	5	6	7	8
1	.10	.19	.16	15	.15	.14	.14	14	.14
	.05	.23	.19	.18	.18	.17	.17	.17	.16
	.01	.31	.26	.24	.23	.23	.22	.22	.22
2	.10		.13	.12	.11	.11	.11	.10	.10
	.05		.16	.14	.14	.13	13	.12	.12
	.01		.21	.19	.18	.17	.17	.16	.16
3	.10			.11	.10	.10	.09	.09	.09
	.05			.13	.12	.11	11	.11	.10
	.01			.17	.16	.15	.15	.14	.14
4	.10				.09	.09	.08	08	.08
	.05				.11	.10	.10	.10	.09
	.01				.15	.14	.13	.13	.13
5	.10					.08	.08	.08	.07
	.05					.10	.09	.09	.09
	.01					.13	.12	.12	.12
6	.10						.07	.07	.07
	.05						.09	.09	.08
	.01						.12	.11	.11
7	.10							.07	.07
	.05							.08	.08
	.01							.11	.11
8	.10								.06
	.05								.08
	.01								.10

m_2	α	m_1 = 15	14	13	12	11	10	9
15	.10	.05						
	.05	.06						
	.01	.07						
14	.10	.05	.05					
	.05	.06	.06					
	.01	.08	.08					
13	.10	.05	.05	.05				
	.05	.06	.06	.06				
	.01	.08	.08	.08				
12	.10	.05	.05	.05	.05			
	.05	.06	.06	.06	.06			
	.01	.08	.08	.08	.08			
11	.10	05	.05	05	.05	.06		
	.05	.06	.06	.06	.06	.07		
	.01	.08	.08	.08	.09	.09		
10	.10	.05	.05	.05	.06	.06	.06	
	.05	.06	.06	.06	.07	.07	.07	
	.01	.08	.08	.09	.09	.09	.09	
9	.10	.05	.06	.06	.06	06	.06	.06
	.05	.07	.07	07	.07	.07	.07	.07
	.01	.09	.09	.09	.09	.09	.09	.10
8	.10	.06	.06	.06	.06	.06	.06	.06
	.05	.07	.07	.07	.07	.07	.07	.08
	.01	.09	09	.09	.09	.09	.10	.10
7	.10	.06	.06	.06	.06	.06	.06	.07
	.05	.07	.07	.07	.07	.07	.08	.08
	.01	.09	.09	.10	.10	.10	.10	.10
6	.10	.06	.06	06	.06	.07	.07	.07
	.05	.07	.08	.08	.08	.08	.08	.08
	.01	.10	.10	.10	.10	.10	11	.11
5	.10	.07	.07	.07	.07	.07	.07	.07
	.05	.08	.08	.08	.08	.08	.08	.09
	.01	.11	.11	.11	.11	.11	11	11
4	.10	.07	.07	07	.07	.08	.08	.08
	.05	.09	.09	.09	.09	.09	.09	.09
	.01	.11	.12	.12	12	12	.12	.12
3	.10	.08	.08	08	.08	.08	09	.09
	.05	.10	.10	.10	10	.10	.10	.10
	.01	13	.13	.13	.13	.13	.13	.14
2	.10	.10	.10	.10	.10	.10	.10	.10
	.05	.12	.12	.12	.12	.12	12	12
	.01	.15	.15	.15	.16	.16	.16	.16
1	.10	.13	.13	.13	.13	.14	.14	.14
	.05	.16	.16	.16	.16	.16	.16	.16
	01	.21	.21	.21	.21	.21	21	.22

TABLE II—*Continued*

Subgroup Size n = 15

m_2	α	m_1 = 1	2	3	4	5	6	7	8
1	.10	.18	.15	.14	.14	.14	.13	.13	.13
	.05	.22	.18	.17	.17	.16	.16	.16	.16
	.01	29	.25	.23	.22	.21	.21	.21	.21
2	.10		.12	.11	.11	.10	.10	.10	.10
	.05		.15	.14	.13	.12	.12	.12	.12
	01		.20	.18	.17	.16	.16	.16	.15
3	.10			.10	.09	.09	.09	.09	.08
	.05			.12	.11	.11	.10	10	.10
	.01			.16	.15	.14	.14	.13	.13
4	.10				.09	.08	.08	.08	.08
	.05				.10	.10	.10	.09	.09
	.01				.14	.13	.13	.12	.12
5	.10					.08	.07	.07	.07
	.05					.09	.09	.09	.08
	.01					12	.12	.11	.11
6	.10						.07	.07	.07
	.05						.08	.08	.08
	.01						.11	.11	.10
7	.10							.07	.06
	.05							.08	.08
	.01							.10	.10
8	.10								.06
	.05								.07
	.01								.10

m_2	α	m_1 = 15	14	13	12	11	10	9
15	.10	.04						
	.05	.05						
	.01	07						
14	.10	.05	.05					
	.05	.05	.06					
	.01	.07	.07					
13	.10	.05	.05	.05				
	.05	06	.06	06				
	.01	.07	07	08				
12	.10	.05	.05	05	05			
	.05	.06	.06	.06	.06			
	.01	07	.08	.08	.08			
11	10	05	05	.05	.05	.05		
	.05	.06	.06	.06	.06	.06		
	.01	.08	08	.08	.08	.08		
10	.10	.05	.05	05	.05	.05	.05	
	.05	.06	.06	.06	.06	.06	.07	
	.01	08	.08	.08	.08	.08	.09	
9	.10	.05	.05	.05	.05	.06	.06	.06
	.05	.06	.06	.06	.06	.07	.07	.07
	.01	.08	.08	.08	.08	.09	.09	.09
8	.10	.05	.05	.06	.06	.06	.06	.06
	.05	.06	.07	.07	.07	.07	.07	.07
	.01	.08	09	.09	09	.09	.09	09
7	.10	.06	.06	.06	.06	.06	.06	.06
	.05	.07	.07	.07	.07	.07	.07	.07
	.01	.09	.09	.09	.09	.09	.10	.10
6	.10	.06	06	.06	.06	.06	.06	.06
	.05	.07	.07	.07	.07	.07	.08	.08
	.01	.09	.09	.10	.10	.10	.10	.10
5	.10	.06	.06	.06	.07	.07	.07	.07
	.05	.08	.08	.08	.08	.08	.08	.08
	.01	.10	.10	.10	.10	.10	.11	.11
4	10	.07	.07	.07	.07	.07	.07	.07
	.05	.08	.08	.08	.08	.09	.09	.09
	.01	.11	.11	.11	.11	.11	.11	.12
3	.10	.08	.08	.08	.08	.08	.08	.08
	.05	.09	.09	.09	.09	.10	.10	.10
	.01	.12	.12	.12	.12	.13	.13	.13
2	.10	.09	.09	.09	.09	.09	.10	.10
	.05	.11	11	.11	.11	.11	.11	.11
	.01	.15	15	.15	.15	.15	.15	.15
1	.10	.13	13	.13	.13	.13	.13	.13
	.05	.15	.15	.15	.15	.15	.15	.16
	.01	.20	.20	20	.20	.20	.20	.20

distribution of G_2 for $\alpha=.10$, .05 and .01; $n=2, 3, \cdots, 15$; m_1 and $m_2=1, 2, \cdots, 15$.

Tables I and II were computed directly from Lord's tables of the percentage points of the distribution of u, making the transformations indicated above. For certain values of m which were not included in the original tables, the corresponding percentage points for u were obtained using a five-point Lagrangian interpolation formula as suggested by Lord [3].

These tests are of the two-tail variety. The significance levels to be used with one-tail tests are $\alpha/2$ In applications of the G-test the usual assumptions are made concerning normality, homogeneity of variances and randomness in the distribution of sample values among the subgroups.

REFERENCES

[1] Daly, J F., "The use of the sample range in an analogue of Student's t-test," *Annals of Mathematical Statistics*, 17 (1946), 71–4.

[2] Lord, E , "Power of the modified t-test (u-test) based on range," *Biometrika*, 37 (1950), 64–77.

[3] Lord, E , "The use of range in place of standard deviation in the t-test," *Biometrika*, 34 (1947), 41–67.

[4] Tippett, L H C., "On the extreme individuals and the range of samples taken from a normal population," *Biometrika*, 17 (1925), 364–87

THE PROPERTIES OF THE MEAN SQUARE SUCCESSIVE DIFFERENCE IN SAMPLES FROM VARIOUS POPULATIONS

P. G. Moore
University College, London and Princeton University

This paper is concerned with the use of the mean square successive difference as a method of estimating variance and testing for homogeneity in samples from various non-normal populations where the order of the observations can be recorded. It is shown that the relative efficiency of such estimates of variance increases as the kurtosis of the population sampled increases and any trend effect is virtually eliminated. The actual distribution varies but can be approximately dealt with by means of a Pearson Type III curve with the correct moments.

1. INTRODUCTION

In all the previous work on the properties and uses of the mean square successive difference in random samples from some parent population the assumption has been made that the sampling is from a normal parent population. In this paper we consider the case of sampling from non-normal populations and discuss the use of the mean square successive difference both as an estimator of the population variance in certain circumstances and also as a test for the independence or dependence of a sequence of observations. Although analytic expressions are not obtained for the distribution, its form when the sampling is from

a) Normal population,
b) Rectangular population,
c) Double exponential population,
d) Pearson Type III population,

is discussed in general terms from the standpoint of moments and suggestions made as to the best method of obtaining a required significance point The relationship between the forms of the distributions of this statistic and the sample squared standard deviation is also considered and finally a transformation which may be of considerable value in obtaining significance points to the generally required accuracy is suggested

2. HISTORICAL

The usual method for estimating the variance of a population on the

basis of a sample is to use the sample variance

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{2.1}$$

where $x_1, x_2 \cdots x_n$ are the sample observations (2.1) is an unbiased estimator and although other estimators such as the mean deviation or sample range are sometimes used it is generally regarded as the basic method Sometimes we have the situation where a gradual shift in the mean takes place as the sampling progresses and yet we still want the actual value of the variance of a single determination. An example occurs in ballistics and weapon testing where it is almost impossible because of atmospheric conditions to get two successive observations under exactly the same conditions yet for the calculation of range tables and so forth it is necessary to know the variance to be expected amongst a group of projectiles fired under exactly similar conditions. Situations will also occur, though, where this does not apply and it is desired to take any gradual shift into the variability measure and thus obtain an over-all picture of the variability A consumer buying large batches of coal would perhaps be more interested in the variability of the whole batch than in small parts of the batch from the point of view of fixing the conditions under which the coal is being used

If instead of using the estimators mentioned above we use some form of estimator based on successive differences, we will considerably reduce any errors due to a trend effect Vallier [15] estimated the dispersion from successive differences and Cranz and Becker [3] used the mean successive difference

$$d = \frac{1}{n-1} \sum_{i=1}^{n-1} | x_{i+1} - x_i | \tag{2.2}$$

to estimate the dispersion where $x_1, x_2 \cdots x_n$ are the observations in their temporal order. Kamat [8] has discussed the statistic d in some detail for the case of samples from a normal population He has derived the first four moments of d and d/s and discussed the distributions of the ratios d/s and d/σ, giving some percentage points The statistic d/s has value in quality control work where it can be used to detect a trend; its value being lower in this case than it would be in sampling from a homogeneous population. Keen and Page [9] have discussed the use of d in this connection and shown its great potentialities due to the simplicity of application.

An alternative statistic is

$$\delta^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_i - x_{i+1})^2 \tag{2.3}$$

where the x's are in temporal order as before. The statistic was investigated by Von Neumann, Kent, Bellinson, and Hart [11] for the case of samples from a normal population and they discuss the form of the distribution of δ^2 and its use as an estimator of variance. The ratio δ^2/s^2 was further investigated by Williams [16], Von Neumann [10], Hart and Von Neumann [7], and Hart [6] The last paper gives some significance points for δ^2/s^2 in samples from a normal population.

Guest [5] has compared the efficiencies of various statistics using mean differences and mean square differences to estimate the population variance. He demonstrates that if the observations are drawn from a normal population the asymptotic efficiencies of the statistics d and δ^2, as the number of observations in the sample gets large, are 60.5 per cent and 66.7 per cent respectively where the sample variance estimate has, of course, an efficiency of 100 per cent.

In this paper we are going to discuss the use of δ^2 to estimate the variance of the parent population and to compare it with two alternative estimators. We will then examine the form of the distribution of δ^2 in samples from various populations and how approximate significance points might be obtained for the distribution

3. MOMENTS OF δ^2

The statistic that we consider is

$$\delta^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2.$$

Let μ_r be the moments ($r=2, 3, \cdots ,$) about the mean of the sampled population, and μ_1' the first moment about the origin

Now

$$\begin{aligned}(n-1)\delta^2 &= \sum_{i=1}^{n-1}(x_{i+1} - x_i)^2 \\ &= \sum_{i=1}^{n-1}(\overline{x_{i+1} - \mu_1'} - \overline{x_i - \mu_1'})^2 \\ &= \sum_{i=1}^{n-1}(X_{i+1} - X_i)^2, \quad \text{where} \quad X_i = x_i - \mu_1' \\ &= 2\sum_{i=1}^{n} X_i^2 - X_1^2 - X_n^2 - 2\sum_{i=1}^{n-1} X_iX_{i+1}. \end{aligned} \tag{3.1}$$

We now find the moments of (3.1) assuming that all the x's are independent. By taking expectations of (3.1) we have

$$(n-1)\mu_1'(\delta^2) = E\left[2\sum_{i=1}^{n} X_i^2 - X_1^2 - X_n^2 - 2\sum_{i=1}^{n-1} X_iX_{i+1}\right] \quad (3.1)$$

$$= 2(n-1)\mu_2.$$

Similarly

$$(n-1)^r\mu_r'(\delta^2) = E\left[2\sum_{i=1}^{n} X_i^2 - X_1^2 - X_n^2 - 2\sum_{i=1}^{n-1} X_iX_{i+1}\right]^r. \quad (3.2)$$

The right hand side of (3.2) may be expanded and expectations taken term by term of the resulting expressions. The crude moments may now be converted into moments about the mean giving

$$(n-1)\mu_1'(\delta^2) = (2n-1)\mu_2,$$

$$(n-1)^2\mu_2(\delta^2) = 2(2n-3)\mu_4 + 2\mu_2^2,$$

$$(n-1)^3\mu_3(\delta^2) = 2(4n-7)\mu_6 + 6(4n-5)\mu_4\mu_2$$

$$-4(8n-11)\mu_2^3 - 8(7n-13)\mu_3^2,$$

$$(n-1)^4\mu_4(\delta^2) = 2(8n15)\mu_8 + 8(16n-27)\mu_6\mu_2 + 2(24n^2+8n-101)\mu_4^2,$$

$$+48(-13n+24)\mu_4\mu_2^2 + (336n-624)\mu_2^4$$

$$+32(-16n+33)\mu_5\mu_3 + 32(25n-154)\mu_3^2\mu_2. \quad (3.3)$$

We can apply a check to these results for the case of samples from a normal population for which we have

$$\mu_4 = 3\mu_2^2; \quad \mu_6 = 15\mu_2^3; \quad \mu_8 = 105\mu_2^4; \quad \mu_3 = \mu_5 = 0;$$

and our moments (3 3) reduce to

$$(n-1)\mu_1'(\delta^2) = 2(n-1)\mu_2,$$

$$(n-1)^2\mu_2(\delta^2) = 4(3n-4)\mu_2^2,$$

$$(n-1)^3\mu_3(\delta^2) = 32(5n-8)\mu_2^3,$$

$$(n-1)^4\mu_4(\delta^2) = 48(9n^2+46n-112)\mu_2^4 \quad (3.4)$$

and these values agree with those given in [11].

At this stage we will give the moments of (2.1) which are used later.

They have been given in a slightly different form by Church [2].

$$\mu_1'(s^2)=\mu_2,$$

$$\mu_2(s^2)=\frac{\mu_4}{n}-\frac{n-3}{n(n-1)}\mu_2^2,$$

$$\mu_3(s^2)=\frac{\mu_6}{n^2}-\frac{2(3n^2-6n+5)}{n^2(n-1)^2}\mu_3^2-\frac{3(n-5)}{n^2(n-1)}\mu_4\mu_2+\frac{2(n^2-12n+15)}{n^2(n-1)^2}\mu_2^3,$$

$$\begin{aligned}\mu_4(s^2)=&\frac{\mu_8}{n^3}-\frac{4(n-7)}{n^3(n-1)}\mu_6\mu_2-\frac{8(3n^2-6n+7)}{n^3(n-1)^2}\mu_5\mu_3\\&+\frac{3n^4-12n^3+42n^2-60n+35}{n^3(n-1)^3}\mu_4^2\\&+\frac{3\mu_4\mu_2^2}{n^3(n-1)^3}\{-2n^4+42n^3-294n^2+630n-420\}\\&+\frac{16\mu_3^2\mu_2}{n^3(n-1)^3}\{6n^3-27n^2+50n-35\}\\&+\frac{\mu_2^4}{n^3(n-1)^3}\{3n^4-27n^3+279n^2-765n+630\}\end{aligned}\tag{3.5}$$

Finally we will consider a third alternative estimator

$$\eta^2=\frac{1}{2m}\sum_{i=1}^{m}(x_{2i}-x_{2i-1})^2,\tag{3.6}$$

where $n=2m$, that is we are considering the case of n even. The moments may be written down straightforwardly as

$$\mu_1'(\eta^2)=\mu_2,$$

$$\mu_2(\eta^2)=\frac{\mu_4}{2m}+\frac{\mu_2^2}{2m},$$

$$\mu_3(\eta^2)=\frac{\mu_6}{4m^2}+\frac{9\mu_4\mu_2}{4m^2}-\frac{5\mu_3^2}{2m^2}-\frac{5\mu_2^3}{2m^2},$$

$$\begin{aligned}\mu_4(\eta^2)=&\frac{\mu_8}{8m^3}+\frac{5\mu_6\mu_2}{2m^3}-\frac{7\mu_5\mu_3}{m^3}+\frac{3m+13}{4m^3}\mu_4^2+\frac{3m-27}{2m^3}\mu_4\mu_2^2\\&+\frac{10}{m^3}\mu_3^2\mu_2+\frac{3m+21}{4m^3}\mu_2^4.\end{aligned}\tag{3.7}$$

4. EFFICIENCY OF ESTIMATORS

In ordinary work we use the sample variance as our estimator of the population variance and this estimator will be taken as the standard although it must be realized that for many populations the sample variance may not be the best possible estimator. Our measure of relative efficiency will be taken as the ratio of the variances of the statistics which are unbiased estimators of the population variance. Hence we have

$$\text{Relative efficiency of} = \frac{\text{Variance of } s^2}{\text{Variance of } \frac{1}{2}\delta^2}$$

$$= \frac{2\{(n-1)^2\mu_4 - (n-1)(n-3)\mu_2^2\}}{n\{(2n-3)\mu_4 + \mu_2^2\}}$$

$$= \frac{2}{n}\,\frac{(n-1)^2\beta_2 - (n-1)(n-3)}{(2n-3)\beta_2 + 1}\cdot \qquad (4.1)$$

where $\beta_2 = \mu_4/\mu_2^2$. We notice that for large n the efficiency tends to the value $(\beta_2 - 1)/\beta_2$ and the greater the value of β_2 the larger the asymptotic relative efficiency of δ^2. Usually as n increases for a given β_2 the efficiency drops away to its limit but for large β_2 it drops to a minimum and then increases slightly to the limiting value. For any given n the efficiency increases with β_2. Table 1 gives some specimen values for this efficiency. We note that β_2 equal to three is the normal population case and that β_2 must always be greater than unity (see, for example, Shohat [14]).

TABLE 1

RELATIVE EFFICIENCY OF δ^2 TO s^2

		β_2							
		1	2	3	4	5	6	10	∞
n	5	0.40	0.64	0.73	0.77	0.80	0.82	0.95	0.91
	10	0.20	0.57	0.69	0.76	0.80	0.82	0.87	0.95
	15	0.13	0.54	0.68	0.75	0.80	0.82	0.88	0.97
	20	0.10	0.53	0.68	0.75	0.80	0.83	0.89	0.98
	25	0.08	0.53	0.68	0.75	0.80	0.83	0.89	0.98
		0	0.50	0.67	0.75	0.80	0.83	0.90	1.00

The statistic η^2 has a relative efficiency, compared with δ^2, given by the expression

$$\text{Relative efficiency of } \eta^2 \text{ to } \delta^2 = \frac{n}{(n-1)^2}\,\frac{(2n-3)\beta_2+1}{2(\beta_2+1)}, \qquad (4.2)$$

where

$$n = 2m.$$

Some specimen values of this efficiency are given in Table 2 We note that as n gets large the efficiency approaches $\beta_2/(\beta_2+1)$ which shows that it never quite becomes as efficient an estimate as δ^2.

TABLE 2

RELATIVE EFFICIENCY OF η^2 TO δ^2, AND IN PARENTHESES, η^2 TO s^2

		β_2				
		1	1.5	3	5	8
n	6	0.60 (0.17)	0.70 (0.36)	0.84 (0.60)	0.92 (0.73)	0.98 (0.82)
	10	0.56 (0.11)	0.65 (0.29)	0.80 (0.56)	0.89 (0.70)	0.94 (0.80)
	20	0.53 (0.05)	0.63 (0.24)	0.78 (0.53)	0.86 (0.68)	0.91 (0.79)
	30	0.52 (0.03)	0.62 (0.23)	0.77 (0.52)	0.85 (0.68)	0.91 (0.79)
	50	0.51 (0.02)	0.61 (0.22)	0.76 (0.51)	0.84 (0.67)	0.90 (0.78)
	100	0.50 (0.01)	0.60 (0.21)	0.76 (0.51)	0.84 (0.67)	0.89 (0.78)

The relative efficiency of η^2 compared with s^2 is given by the expression

$$\text{Relative efficiency of } \eta^2 \text{ to } s^2 = \frac{(n-1)\beta_2-(n-3)}{(n-1)(\beta_2+1)} \qquad (4.3)$$

and the figures for this are given in parentheses in Table 2

In the particular case of $\beta_2=1$ which corresponds to a frequency distribution with just two equal ordinates it is of interest to note that s^2 becomes a "super efficient" estimate since the variance reduces to

$$\mu_2(s^2) = \frac{2}{n(n-1)}\,\mu_2{}^2$$

and thus decreases in the order $1/n^2$ instead of the usual $1/n$

Generally the relative efficiency of δ^2 increases with β_2 and similarly for η^2. The relative efficiency of η^2 also catches up that of δ^2 as β_2 in-

creases. Hence the use of δ^2 will be of greatest value when the parent population is of leptokurtic form. We notice that the skewness of the parent population does not affect our results in any way and only the kurtosis matters. We also see that the size of the sample has remarkably little effect on the relative efficiencies, except for very low values of β_2, and the efficiencies are dependent almost entirely on the β_2 of the parent population.

5 BIAS OF ESTIMATORS

To examine the question of bias we use a method similar to that utilized by Kamat [8] in the case of the statistic d Let the mean of the sampled population be θ_i when the ith observation is taken and the standard deviation be σ throughout Let $\Delta\theta_i/\sigma = (\theta_{i+1} - \theta_i)/\sigma$ be small so that the third and higher powers may be neglected Then by following a similar method to that used in Section 3, taking, for example

$$(n-1)\delta^2 = \sum_{i=1}^{n-1} \left\{ (x_{i+1} - \theta_{i+1}) - (x_i - \theta_i) + (\theta_{i+1} - \theta_i) \right\}^2$$

we obtain the following results, where

$$\bar{\theta} = \frac{1}{n} \sum_{i=1}^{n} \theta_i,$$

$$\mu_1'(\tfrac{1}{2}\delta^2) = \mu_2 \left\{ 1 + \frac{\sum_{i=1}^{n-1} (\Delta\theta_i)^2}{2(n-1)\mu_2} \right\}, \tag{5.1}$$

$$\mu_2(\tfrac{1}{2}\delta^2) = \frac{\mu_2^2}{2(n-1)^2} \left\{ (2n-3)\beta_2 + 1 + \frac{4}{\mu_2} \sum_{i=1}^{n-1} \left[(\Delta\theta_i)^2 - (\Delta\theta_i)(\Delta\theta_{i+1}) \right] + \frac{2\mu_3}{\mu_2^2} (\Delta\theta_{n-1} - \Delta\theta_1) \right\}. \tag{5.2}$$

The last term will be zero for a linear trend with observations taken at equal intervals and can usually be safely ignored. For η^2 we have

$$\mu_1'(\eta^2) = \mu_2 + \frac{1}{2m} \sum_{i=1}^{m} (\Delta\theta_{2i-1})^2, \tag{5.3}$$

$$\mu_2(\eta^2) = \frac{\mu_2^2}{2m} \left\{ \beta_2 + 1 + \frac{1}{2m\mu_2} \sum_{i=1}^{m} \Delta\theta_{2i-1}^2 \right\}, \tag{5.4}$$

and for s^2, we have

$$\mu_1'(s^2) = \mu_2\left\{1 + \frac{\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2}{(n-1)\mu_2}\right\}, \tag{5.5}$$

$$\mu_2(s^2) = \mu_2^2\left\{\frac{\beta_2}{n} - \frac{n-3}{n(n-1)} + \frac{4}{(n-1)^2\mu_2}\sum_{i=1}^{n}(\theta_i - \bar{\theta})^2\right\}. \tag{5.6}$$

Equations (5.1) and (5.5) are given by Kamat [8] and in the cases where the drawings come from a normal population (5.2) and (5.6) reduce to the values that Kamat gives.

We will take two specific cases as illustrations of these general formulas.

Example 1

$$\theta_i = \mu_1' + (0.05)i\sigma \qquad i = 1, 2, \cdots, n = 20$$

The percentage bias in the estimate $\delta^2 = 0.125$
The percentage bias in the estimate $\eta^2 = 0.125$
The percentage bias in the estimate $s^2 = 8.750$

Example 2

$$\theta_i = \mu_1' + (0.02)i\sigma \qquad \text{for } i = 1, 2, \cdots, n_1 = 10$$
$$= \mu_1' + 0.2 - 0.03(i - 10)\sigma \qquad \text{for } i = 11, 12, \cdots, n = 20$$

This is a case where the mean increases slowly linearly and then decreases linearly at a different rate.

The percentage bias in the estimate $\delta^2 = 0.033$
The percentage bias in the estimate $\eta^2 = 0.032$
The percentage bias in the estimate $s^2 = 0.728$

We notice how close the bias in δ^2 and η^2 come. In fact for a trend that is exactly linear all the time they give the same percentage bias and as δ^2 has the lower variance it is clearly better to use it. s^2 has a very much larger bias. In example 2 the bias is 23 times as large as that of δ^2 and 70 times as large in example 1. This disparity is mainly due to a certain amount of averaging out in s^2 in example 2 due to the trend increasing and then decreasing.

It is also of interest to see the effect that these slow shifts have on the variance of the estimators. First we make a few general remarks. The variances do not depend in any way, to the order of approximation con-

sidered, on the skewness of the populations sampled but do depend on the amount of kurtosis present. If we imagine taking a sample of size n with a fixed set of values θ_i, then for a whole series of populations with the same variance the populations with the lower values of β_2 will produce the lower sample variances of the estimators of variance If we consider example 1 again and imagine sampling from the three populations

a) Rectangular $\beta_2=1\ 8$
b) Normal $\beta_2=3$
c) χ^2 type $\beta_2=6$

all three populations having the same variances then the percentage change in the variance of the estimators over the variances that would occur with no trend present are as given in Table 3

TABLE 3

PERCENTAGE CHANGES IN VARIANCES OF ESTIMATORS

Estimator	Rectangular	Normal	χ^2 type
δ^2	0	0	0
η^2	0 045	0 031	0.018
s^2	40.698	17 500	7 216

From this table we see that δ^2 and η^2 give a negligible change, in fact for a linear change δ^2 gives zero, whilst s^2 gives a huge difference which diminishes as β_2 increases We must emphasize that this variance is *that of the three statistics in repeated samples drawn from the populations* specified where the linear trend of example 1 holds

6 PARENT POPULATION NORMAL

We turn now to a consideration in this and the succeeding three sections of the form of the distribution of δ^2 in samples from various types of population and how to obtain quick and approximate significance points First we consider the parent population to be normal In this case from (3 4) we have

$$\beta_1 = \frac{16(5n-8)^2}{(3n-4)^3}, \qquad \beta_2 = \frac{3(9n^2+46n-112)}{(3n-4)^2}$$

and in the first three columns of Table 4 the values of β_1 and β_2 are given for a series of values of n We notice that the values tend to the normal curve values as n increases Plotting them in a chart of (β_1, β_2) field as in Figure 1 we find that they fall in the Type VI region between

TABLE 4

VALUES OF β_1 AND β_2 IN SAMPLES FROM NORMAL POPULATION

n	β_1	β_2	1st approximation			2nd approximation		
			ν	β_1	β_2	ν	β_1	β_2
10	1 6058	5 5385	6 2308	1 2839	4 9258	4 9819	1.6058	5 4087
20	0 7711	4 2168	12 8928	0 6205	3 9307	10 3743	0 7711	4 1566
30	0 5072	3 7999	19 5581	0 4090	3.6135	15 7721	0.5072	3 7608
40	0 3779	3 5957	26 2241	0 3051	3 4077	21.1710	0 3779	3 5668
50	0 3011	3 4746	32 8900	0 2432	3.3648	26 5704	0 3011	3 4516
75	0 2007	3 3146	49 5566	0 1614	3 2421	40 0696	0 1997	3.2995
100	0 1493	3 2353	66 2230	0 1208	0 1812	53 5692	0 1493	3 2239

the bounding lines of the Type III and Type V curves and rather nearer to the former than the latter If the (β_1, β_2) points for the log normal distribution are also put on the chart, it is found that our values are above them as well. To obtain approximate significance points we may fit a frequency curve whose moments agree with those of the statistics concerned. Three methods suggest themselves, namely to use a Type III or χ^2 distribution, a Type VI or a log normal distribution. Since the

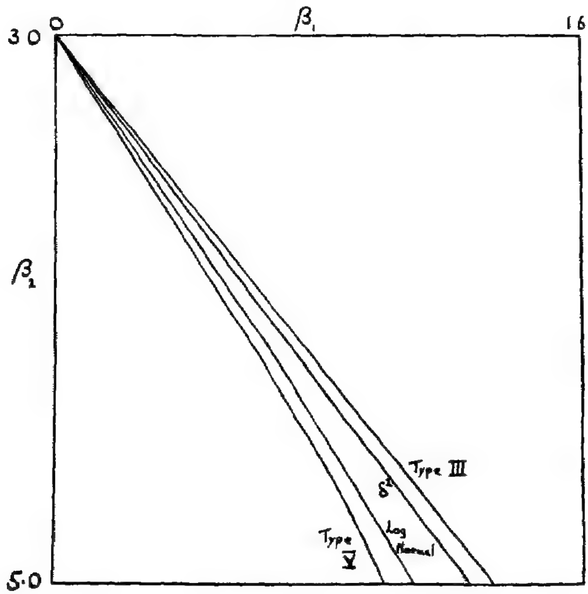


FIGURE 1

object is to be able to obtain a quick method of getting significance levels, the first method which leads up to an extensively tabulated function would appear to be very much more attractive than the other two, both of which require a three moment fit and, for the Type VI, tables of the incomplete beta-function.

To use a Type III approximation we take δ^2/σ^2 to be distributed in the same way as $c_1\chi^2/\nu$ where the χ^2 has ν degrees of freedom. If we now make the first two moments agree we find

$$c_1 = 2, \qquad \nu = 4(n-1)^2/[(2n-3)\beta_2 + 1]. \tag{6.1}$$

Using these values we give in Table 4 under the heading 1st approximation the resulting values for ν and hence for β_1 and β_2 since for a χ^2 distribution

$$\beta_1 = 8/\nu, \qquad \beta_2 = 3 + 12/\nu \tag{6.2}$$

We see that for all the values of n the fitted approximate (β_1, β_2) point is nearer the normal point, as regards both β_1 and β_2, than before. The effect of using the approximate fitted χ^2 instead of a complete Type VI solution is illustrated for the case of $n=50$ in Table 5 under the 1st approximation where the Type VI values are taken from [11]. The

TABLE 5

PROBABILITY $(\delta^2/\sigma^2) < \xi$ FOR $n=50$

ξ	Type VI	1st approx	2nd approx	Normal
0 30	0 00000	0 00001	0 00000	0 00118
0 75	0 00031	0 00049	0 00023	0.00563
1 00	0 00674	0 00797	0.00619	0 02129
1 25	0 04393	0 04825	0.04407	0 06418

normal column is obtained by assuming δ^2 to be normally distributed with the mean and variance given by (3 4). Our approximation slightly over-estimates the probabilities, as measured by the Type VI which is most likely to be nearest the true value, but not nearly so much as the normal approximation and it would certainly be accurate enough for most purposes The approximation can be improved by using a fitting of the form $c_1\chi^2/\nu + c_2$ for δ^2/σ^2. To fix the constants we must now make three moments agree giving

$$c_1 = \frac{(3n-4)^2}{(n-1)(5n-8)}, \qquad c_2 = \frac{n(n-2)}{(n-1)(5n-8)},$$
$$\nu = \frac{(3n-4)^3}{2(5n-8)}. \tag{6.3}$$

The new values of ν and the corresponding values for β_1 and β_2 are given under the 2nd approximation in Table 4. In this case the β_1's agree due to the three moment fit and the values of β_2 are slightly lower than the true values. In Table 5 under the heading "2nd approximation" the values of the probabilities obtained using this second method are given. This second approximation makes the start at a distance c_2 above zero instead of zero which must be the true start although for all practical purposes the start will be above zero We also note that in this particular case η^2 will be distributed as χ^2 with m degrees of freedom since each pair of values gives rise to the square of a normal variate which is distributed as χ^2 with one degree of freedom

Finally in Table 6 we give some approximate upper 5 per cent points for δ^2/σ^2 calculated by four different methods. The χ^2 approximations and the normal approximation are those we have just described and the Pearson Type VI curve uses the value of the upper 5 per cent points of the Pearson curves given by Pearson and Merrington [12] The case of n equal to 10 fell outside the latter table and was computed directly from a fitted Type VI curve From these results it seems apparent that for any reasonable size n the first χ^2 approximation will provide an adequate significance point even if the normal distribution is not adequate

TABLE 6

APPROXIMATE UPPER 5% POINTS FOR δ^2/σ^2

Value of n	10	20	30	40	50	75	100
χ^2 1st approx	4 15	3 45	3 15	2 99	2 88	2 70	2 61
χ^2 2nd approx	4 17	3 46	3 16	2 99	2 88	2 71	2 61
Normal	4 22	3 54	3 25	3 08	2.96	2 78	2.68
Pearson curve	4.17	3 46	3 16	2 99	2 88	2.71	2 61
Log Approx. (section 11)	4 33	3 54	3 22	3 03	2 91	2 73	2 62

7 PLATYKURTIC PARENT POPULATION

If the parent population be taken as rectangular with

$$p(x) = 1, \quad 0 \leqq x \leqq 1$$
$$= 0, \quad \text{elsewhere} \tag{7.1}$$

then all the odd moments are zero and the even moments have the values

$$\mu_2 = \frac{1}{12}, \qquad \mu_4 = \frac{1}{80}, \qquad \mu_6 = \frac{1}{448}, \qquad \mu_8 = \frac{1}{2304}.$$

The value of β_1 is zero and β_2 is 1 8 giving a platykurtic distribution. In practice we would not estimate variance by the sample variance in this case but it is used here as an illustration of a platykurtic distribution. Table 7 gives the values of β_1 and β_2 of δ^2 for the same sample sizes as before. The values now are nearer the normal point than they were for samples of the same size from a normal population and we find that

TABLE 7

MOMENTAL RATIOS FOR SAMPLES FROM RECTANGULAR POPULATION

n	β_1	β_2	1st approximation			2nd approximation		
			ν	β_1	β_2	ν	β_1	β_2
10	0 5546	3 7282	10 2532	0 7802	4 1703	14 4248	0 5546	3 8219
20	0 2673	3 3137	21 3609	0 3745	3 5617	29 9065	0 2673	3 4011
30	0 1760	3 1994	32 4710	0 2464	3 3696	45 4545	0 1760	3 2640
40	0 1318	3 1460	43 5692	0 1836	3 2754	60 6980	0 1318	3 1977
50	0 1048	3 1152	54 6925	0 1463	3 2195	76 3358	0 1048	3 1572
75	0 0694	3 0754	82 4699	0 0970	3.1455	115 2738	0 0694	3 1041
100	0 0520	3.0560	110 2474	0 0726	3 1089	153 8640	0.0520	3 0780

they actually fall in the Type I region of the (β_1, β_2) plane The Type I curve is very tedious to fit having four constants to be fixed and a χ^2 may once again be a suitable approximation in that it has the range starting at zero and is near the actual momental ratios. Using the same procedure as before we find

$$c_1 = 2, \qquad \nu = (n-1)^2/(0\,9n - 1\,1), \tag{7.2}$$

and the fitted points are given in the table A gain in accuracy may once more be obtained by making the start slightly positive, the values being given under the second approximation From this it seems that χ^2 can be used as a reasonable representation of the distribution

8. LEPTOKURTIC PARENT POPULATION

As an example of this kind of parent population we consider the first law of Laplace or the double exponential distribution where

$$p(x) = \tfrac{1}{2}e^{-|x|}, \qquad -\infty < x < \infty \tag{8.1}$$

It is easy to show for this distribution that $\mu_{2r} = (2r)!$, whilst the odd moments are all zero. Hence we have

$$\mu_2 = 2, \qquad \mu_4 = 24, \qquad \mu_6 = 720, \qquad \mu_8 = 40320,$$

giving $\beta_2 = 6$ which indicates a leptokurtic distribution. Substituting

TABLE 8

MOMENTAL RATIO IN SAMPLES FROM DOUBLE EXPONENTIAL POPULATION

n	β_1	β_2	1st approximation		
			ν	β_1	β_2
10	5 4840	13 3579	3 1456	2.5432	6.8148
[illegible]	2 6193	7 9196	6 4753	1 2355	4 8533
30	1 7200	6 2249	9 8076	0 8157	4 2235
40	1 2803	5 3985	13 1404	0 6089	3 9134
50	1 0196	4 9092	16 4734	0 4856	3 7284
75	0.6757	4 2644	24 8063	0.3225	3 4838
100	0 5052	3.9451	33 1395	0.2414	3 3621

these moments in (3 4) we get the β_1, β_2 values for δ^2 and they are given in Table 8 Once again they fall in the Pearson Type VI region, this time very close to the log-normal line.

The β_1 and β_2 points for the first kind of χ^2 approximation are given. For the second kind of approximation the values of β_1 would agree but the relative accuracy of β_2 would not be as good as it was in the case of sampling from a normal population because these points all fall below the locus of points for δ^2 in samples from a normal curve

It thus seems that as the kurtosis of the parent population rises the resultant locus of β_1, β_2 points for δ^2 pivots approximately on the normal point and swings from the Type I region through the Type III line into the Type VI region

9 SKEW PARENT POPULATION

So far we have considered parent populations with varying amounts of kurtosis but they have all had zero skewness We now take a case where there is both skewness and kurtosis present

The particular distribution that has been used is that of a χ^2 variate with ν degrees of freedom For this distribution we have

$$\mu_1' = \nu, \qquad \mu_2 = 2\nu, \qquad \mu_3 = 8\nu, \qquad \mu_4 = 12\nu(\nu + 4),$$

$$\mu_5 = 32\nu(5\nu + 12), \qquad \mu_6 = 40\nu(3\nu^2 + 52\nu + 96),$$

$$\mu_8 = 16\nu(105\nu^3 + 4760\nu^2 + 29{,}232\nu + 40{,}320)$$

As illustrations we take two cases (i) $\nu = 4$ and (ii) $\nu = 12$ For the former $\beta_1 = 2$, $\beta_2 = 6$ whilst for the latter $\beta_1 = 2/3$, $\beta_2 = 4$ In Table 9 the resulting values for β_1 and β_2 are given

It will be seen that they hug the Type III line fairly closely for large n and the Type V for low n being all the while in the Type VI region which accords with the statement made earlier. In this case the appropriate χ^2 approximation would, for large n, give us quite a reasonable figure although for small n it would be better to fit a Type V or inverted χ^2 with appropriate moments. It is of interest to note how the difference between the upper 5 per cent points for a Type III and Type V curve with the same β_1 vary For a β_1 of 0.05 the difference is 0.02, for β_1 equal to 1 it is 0.04. These figures are in terms of the standard deviation of the population as unit The means of our distributions are always 2 so that unless the standard deviation is large the error is likely to be small in taking a Type III approximation with correct β_1 when the true distribution actually lies between the Type III and Type V boundaries. It must also be remembered that in some cases the distribution lies above the Type III line so that by using that as the approximation we are playing for safety and should be nearer the true figure in a large number of cases

TABLE 9

MOMENTAL RATIOS FOR SAMPLES FROM TYPE III POPULATION

n	$\nu=4$		$\nu=12$	
	β_1	β_2	β_1	β_2
10	6 0942	14 8563	3 5258	8 6772
20	2 9334	8 6259	1.6815	5.7279
30	1 9255	6 7097	1 1037	4 7943
40	1.4330	5 7605	0 8214	4 3367
50	1.1411	5 1981	0 6541	4 0650
75	0 7561	4 4563	0 4334	3 7062
100	0 5653	4.0888	0 3241	3 5282

10 DISTRIBUTION OF s^2

In order to make some comparisons we now make a few general remarks about the form of the distribution of s^2 whose moments were given in (3 6). For the case of samples of size n from a normal population it is known that the sample variance is distributed exactly as $\chi^2\sigma^2/(n-1)$ where χ^2 has $(n-1)$ degrees of freedom For the cases of sampling from a Type III population with varying amounts of skewness and kurtosis the momental ratios are given in Table 10 On comparing the sets of momental ratios with those of δ^2 for the correspond-

TABLE 10
MOMENTAL RATIOS

n	Normal pop		Type III population			
	$\beta_1=0$	$\beta_2=3$	$\beta_1=2$	$\beta_2=6$	$\beta_1=\frac{2}{3}$	$\beta_2=4$
	β_1	β_2	β_1	β_2	β_1	β_2
10	0.89	4.33	5.41	15.52	2.55	7.49
20	0.42	3.63	2.70	9.45	1.26	5.31
30	0.28	3.41	1.80	7.32	0.84	4.54
40	0.21	3.31	1.34	6.23	0.62	4.16
50	0.16	3.25	1.08	5.58	0.51	3.95
75	0.11	3.16	0.71	4.69	0.32	3.57
100	0.08	3.12	0.53	4.19	0.25	3.41

ing populations we notice a marked similarity of the way they progress and of the order of magnitude especially for the case of a Type III parent population.

11. TRANSFORMATION OF δ^2

A fruitful idea would be to find some form of transformation for δ^2 so that it will make the test more robust in the sense that the significance points depend very little on the actual form of the distribution of the original observations. One obvious possibility is to use the logarithm of the quantity δ^2. It has not proved possible so far to obtain the distribution of log δ^2 but we know the moments of log s^2 and can appeal to the similarity of δ^2 and s^2. These moments may be obtained directly from David [4] in terms of the population cumulants In some instances it is easier to apply than in this way whereas in other cases it is easier to work in terms of the population moments For the case of samples from a normal distribution Bartlett and Kendall [1] have given the exact values in terms of the derivatives of the logarithm of the gamma function, tabulated the moments up to n equal to 20 and given working approximations for n over 20 The values given in Table 11 are taken from [1] for n equal to 10 and 20 whilst for n over 20 they are obtained from a more accurate approximation than that given in [1]. For the case of sampling from a non-normal population we give in Table 11 the values of the momental ratios for the case of samples from the two Type III populations considered earlier. The values for n equal to 10 are not given as the approximations proved unreliable for n as low as 10. In all cases there is a considerable movement, for any given n, towards normality and it would therefore seem feasible to use the log-

TABLE 11

MOMENTAL RATIOS OF Log s^2

n	Normal pop		Type III population			
			$\beta_1=4$	$\beta_2=6$	$\beta_1=\frac{2}{3}$	$\beta_2=4$
	β_1	β_2	β_1	β_2	β_1	β_2
10	0 2200	3 4370	—	—	—	—
20	0 1050	3 2100	0.2728	3 4992	0.0572	3 2280
30	0 0710	3 1287	0 0770	3 3944	0 0201	3 1528
40	0 0525	3 0974	0 0242	3 3058	0 0098	3.1133
50	0 0416	3 0784	0 0084	3 2453	0.0058	3 0889
75	0 0274	3 0689	0 0005	3 1617	0 0022	3 0566
100	0 0204	3 0396	0 0000	3.1176	0 0011	3.0444

arithm for a quick test of significance using normal curve factors Since there is such a marked similarity between s^2 and δ^2, it would seem reasonable that the logarithmic transformation would be of use in the case of the latter as well.

To carry out a test of significance of δ^2 in this way we would take

$$L = \log_e \delta^2,$$

$$\mu_1'(L) = \log_e 2\mu_2 - \frac{(2n-3)\beta_2 + 1}{4(n-1)^2},$$

$$\sigma(L) \doteqdot \sqrt{2(2n-3)\beta_2 + 2}/2(n-1), \tag{11.1}$$

where β_2 and μ_2 refer to the sampled population. Then the ratio

$$(L - \mu_1'(L))/\sigma(L)$$

should be referred to the normal probability scale The first moment is accurate to order $1/n$ only and the second moment is obtained by using the relation that if $y=f(x)$ then

$$\frac{\sigma_y}{\sigma_y} \doteqdot \left|\frac{dy}{dx}\right|_{x=\xi},$$

ξ being the mean value of x.

In Table 6 the last row gives the upper 5 per cent points for s^2 using (11.1) and it will be seen that they approach the Pearson curve values more quickly than the normal approximation. If more terms were taken in the expressions given in (11.1), the approach would be very much quicker but the test then loses its simplicity and thus its merit.

12. ILLUSTRATIONS

In this section we will consider four examples that illustrate some of the points that have been made in the preceding sections although these by no means exhaust the possibilities

Example (*i*). Counts are being made of the intervals between radioactive particle emissions and it is desired to estimate the variability The first forty intervals observed are:

5.2, 6 5, 6.7, 2.3, 1.4, 0 1, 4 6, 3 2, 3 8, 0 6, 1.2, 17 8, 0 9, 0.3, 1 2, 2.5, 0.9, 0 1, 5.2, 7.0, 1 5, 2 4, 0 1, 4 5, 0.7, 3 0, 9 3, 1 7, 4 8, 3.9, 5 1, 0 5, 0 8, 1 0, 0 2, 3 9, 2 6, 4 7. 4 1, 4 0

Using the successive difference technique we get 11 28 as our estimate of the population variance instead of 10 43 using s^2 Any hesitation we might have in using the successive difference method of estimation would be diminished when we realize that these counts come from an extremely leptokurtic population, one in which the β_2 is approaching a figure of about 9 Hence this method of estimation is 90 per cent efficient when compared with the usual variance estimate and added to this it is very much quicker in calculation time than the usual variance estimate. There is the added advantage that further observations may be immediately incorporated as they come along without any need for fresh calculations of means and so forth.

Example (*ii*) In Table 12 below is given a sample of 40 individuals drawn from a population believed to be normal and an estimate of the variance of that population is required The values, in order of occurrence, go down the columns.

From these figures we obtain straightforwardly

$$\begin{aligned} &\text{Mean} &&= 11.378 \\ &\text{Variance } (s^2) &&= 4\ 586 \\ &\tfrac{1}{2}\delta^2 &&= 3.564 \end{aligned}$$

TABLE 12

RANDOM SAMPLE FROM NORMAL POPULATION WITH TREND

10.52	14.69	7 10	9 73	13 66
7 80	11.52	9.00	12.25	13 85
6 86	12 68	14.55	8 65	16.20
9.82	13.45	13.34	13.84	11.80
10.04	14.57	9.56	11.23	11 37
9.83	9 94	9.82	12 38	13 79
8.71	13.66	11.57	13.42	9 93
10.85	10.73	13.64	11.57	11.19

and we notice that the latter gives a very much smaller estimate than the usual variance estimate. In actual fact the above series was obtained artificially as follows. Using the tables of random normal deviates due to Wold [17] a sample of size 40 was drawn with mean 10 and variance 3 133. To this series was added a trend which increased the rth value by an amount 0 05r Thus over the whole series of 40 observations there was a trend of 1 13 times the standard deviation or 0 03 of a standard deviation between each successive observation. From the expressions given in Section 5 we find

$$\mu_1'(\tfrac{1}{2}\delta^2) = 3.351, \quad \mu_2(\tfrac{1}{2}\delta^2) = 0.856,$$
$$\mu_1'(s^2) = 3.641, \quad \mu_2(s^2) = 0.676.$$

Further whilst there is no increase in $\mu_2(\frac{1}{2}\delta^2)$ as against the case where there is no trend present, there is a percentage increase of 17 per cent in $\mu_2(s^2)$ under the same conditions For μ_1', δ^2 has a percentage increase of only 0 04 per cent as against 8 69 per cent in the case of s^2. The estimate using δ^2 in this case is still high but it is well within the sampling fluctuations we can expect to find in such a sample.

Example (iii). In Table 13 we give the resistance in megohms of 50 pieces of electrical insulating material. All the figures have been divided by 5 and given an arbitrary origin of 900 for the purposes of clarifying the discussion The data is taken from Shewhart [13] and the order of occurrence of the items is down the columns.

Let us suppose that some form of quality control chart was being put into operation based on these figures. We find

$$\text{Mean} = 1.70$$
$$\text{Variance } (s^2) = 9{,}975\ 00$$

and hence

$$\text{Standard deviation} = 99\ 87.$$

TABLE 13

RESISTANCES OF MATERIALS (ARBITRARY UNITS AND ORIGIN)

109	−171	27	28	58
− 30	−148	44	79	− 32
− 30	−240	62	58	79
−105	−163	13	69	250
− 42	−207	−18	40	48
− 14	140	−87	20	100
− 3	120	13	−78	79
− 43	27	138	−18	− 49
−104	120	45	−64	− 66
−115	190	28	58	−130

Plotting the figures on a graph, however, shows a marked pattern and there appears to be a change in mean value over the period If we now compute an estimate of the population variance using δ^2 we obtain the estimate 4256 3 giving a standard deviation of 65 24 This difference would have a profound effect on any control limits that we might put on the chart For instance suppose that we decided to use the means of four items as our criterion then using the 99 per cent probability factors from the normal tables our limits come out to be

Using s^2	Upper 130 3	Lower -126.9
Using δ^2	Upper 85 7	Lower $-$ 82.3.

Further the means of the first twelve groups of four from Table 13 are

-14, -25 5, -134 5, -117 5, 114 25, 36.5, 11 5, 45, 46 75, -25 5, 88.75, 44 5

If we use the control limits obtained from the sample variance, we find one mean only, namely -134 5, outside the limits. The other limits however give four means outside, namely -134 5, -117 5, 114.25 and 88.75. Thus a very different picture arises and a decision must be made as to which should be regarded as the appropriate model. In practice we would like to have a very long series of observations known to be in statistical "control" before computing a variance for purposes o control limits. Until this stage is reached, however, it seems that we can arrive at very different conclusions purely by considering our material from different viewpoints and in general it would be more usual to adopt the cautious view, either by using some estimator based on the dispersion in small sub groups or else by using a successive difference method.

Example (*iv*). From a table of random numbers the following sequence of 100 two digit numbers was obtained the order being across the rows.

07, 94, 31, 40, 57, 79, 06, 72, 99, 23, 78, 61, 12, 39 63, 76, 74, 18, 69, 29, 82, 08, 30, 36, 82, 38, 54, 81, 62, 15, 59, 92, 27, 29 28, 97, 30, 35, 21, 39, 09, 91, 37, 55, 60, 19, 04, 67, 47, 74

These numbers form part of a table which has already been tested for the frequency of occurrence of the various digits and we are interested here in testing whether the order of occurrence is random or not. We find δ^2/σ^2 to be equal to 2 3155, σ^2 being calculated theoretically for this particular case of a retangular distribution. The significance level corresponding to this value for δ^2/σ^2 may be obtained in one of three ways. First we may use the χ^2 approximation given in Section 6 assuming the underlying population to be normal. Secondly we may use the

modification given in (7 2) which assumes an underlying rectangular population and finally we may modify (7 2) with a three moment fit. These three methods give rise to the following levels of significance

Normal basis	0.250
Rectangular basis (Eqn. 7.2)	0 201
Rectangular basis (3 moment fit)	0 202

and thus although not significant by any method the use of the correct form of distribution brings about a considerable change in the significance level This difference can of course prove very important. For example a page was picked out of a telephone directory at random and two digit numbers formed by taking the middle two of the four digits in each of the numbers Fifty such numbers were formed and δ^2/σ^2 came to be 1 2847 where σ^2 was calculated, as before, on the basis of a rectangular distribution.

The significance levels corresponding to this result are 0 061 using the normal curve basis and 0 021 using the rectangular basis and a very different conclusion would no doubt be drawn from the latter figure that would be completely overlooked by the former.

13. CONCLUSIONS

The mean square successive difference, δ^2, can be used as a measure of dispersion in situations where it is desired to eliminate any effects that may cause a gradual drift in the mean value of the sampled population Calculations show that although δ^2 will never be as efficient an estimator as the sample variance, s^2, in situations where no drift occurs, nevertheless as the kurtosis of the sampled population increases the relative efficiency of δ^2 increases very rapidly When some trend does exist, the bias in is negligible compared with that of s^2 and may hardly change at all.

The actual distribution of δ^2 is also of interest in that it may be used where we desire to make a test for homogeneity. Here we find that the form of the parent population does change the form of the distribution of δ^2 to quite a large extent but that it is always somewhere in the neighbourhood of a Pearson Type III curve A suitable approximation to the four moment type of solution can be made for most practical purposes by using a χ^2 distribution with the first two moments in agreement.

We then illustrate how the distribution of s^2, judged by its momental ratios, has a form very similar to that of δ^2. A logarithmic transformation is of use for the former in that it brings it to approximate normality for lower values of sample size and this suggests that a similar transformation would be of use with δ^2 to give a quick significance test.

Finally the methods described are illustrated on a number of examples. These bring out the differences that exist between the properties of estimating variance by successive differences from normal and non-normal populations and also how the significance points for δ^2 vary according to the basic underlying distribution.

REFERENCES

[1] Bartlett, M. S., and Kendall, D. G., "The statistical analysis of variance—heterogeneity and the logarithmic transformation," *Supplement to the Journal of the Royal Statistical Society*, 8 (1946), 128–38

[2] Church, A. E. R., "On the moments of the distribution of squared standard deviations for samples of N drawn from an indefinitely large population," *Biometrika*, 17 (1925), 79–83

[3] Cranz, C., and Becker, K., *Exterior Ballistics* London: His Majesty's Stationery Office, 1921.

[4] David, F. N., "The moments of the z and F distributions," *Biometrika*, 36 (1949), 394–403.

[5] Guest, P. G., "The estimation of the standard error from successive finite differences," *Journal of the Royal Statistical Society, Series B*, 8 (1951), 233–37

[6] Hart, B. I., "Significance levels for the ratio of the mean square successive difference to the variance," *Annals of Mathematical Statistics*, 13 (1942), 445–47.

[7] Hart, B. I., and von Neumann, J., "Tabulation of the probabilities for the ratio of the mean square successive difference to the variance," *Annals of Mathematical Statistics*, 13 (1942), 207–14

[8] Kamat, A. R., "On the mean successive difference and its ratio to the root mean square," *Biometrika*, 40 (1953), 116–27.

[9] Keen, J., and Page, D. J., "Estimating variability from the differences between successive readings," *Applied Statistics*, 2 (1953), 13–23.

[10] von Neumann, J., "Distribution of the ratio of the mean square successive difference to the variance," *Annals of Mathematical Statistics*, 12 (1941) 367–95.

[11] von Neumann, J., Kent, R. H., Bellinson, H. R., and Hart, B. I., "The mean square successive difference," *Annals of Mathematical Statistics*, 12 (1941), 153–62.

[12] Pearson, E. S., and Merrington, M., "Tables of the 0.5% and 5% points of Pearson curves (with argument β_1 and β_2) expressed in standard measure," *Biometrika*, 38 (1951), 4–10.

[13] Shewhart, W. A., *Economic Control of Quality of Manufactured Product*. New York. D. Van Nostrand Co., Inc., 1931.

[14] Shohat, J., "Inequalities for moments of frequency functions and for various statistical constants," *Biometrika*, 21 (1929), 361–70

[15] Vallier, E., *Balistique Experimentale*. Paris: Berger-Levrault and Cie., 1894.

[16] Williams, J. D., "Moments of the ratio of the mean square successive difference to the mean square difference in samples from a normal universe," *Annals of Mathematical Statistics*, 12 (1941), 239–41.

[17] Wold, H., *Random Normal Deviates*, Tracts for Computers XXV. Cambridge: Cambridge University Press, 1948.

SAMPLING PLANS FOR INSPECTION BY VARIABLES*

GERALD J LIEBERMAN AND GEORGE J. RESNIKOFF
Stanford University

I. INTRODUCTION

IN LOT-BY-LOT acceptance sampling by attributes each item of a sample drawn from a lot of manufactured items is classified simply as defective or non-defective A random sample is drawn from the lot and the lot is either accepted or rejected depending solely upon the number of defectives in the sample. Inspection procedures by variables are based on the measurement of a variable quality characteristic, and the decision to accept or reject the lot is a function of these measurements (as opposed to the number of defectives). Variables inspection is applicable whenever the testing of individual items involves measurement and the form of the distribution is known. Since inspection by variables, when it is applicable, makes greater use of the information concerning the lot than does inspection by attributes, variables plans require smaller sample sizes than attributes plans furnishing the same protection. Variables sampling plans pertain to a single quality characteristic, and it is usually assumed that measurements of this quality characteristic are independent, identically distributed normal random variables. Such an assumption will be made throughout the paper.

For the purposes of this paper, sampling inspection by variables is divided into three categories, known standard deviation plans, unknown standard deviation plans, and average range plans Known standard deviation plans are based upon the sample mean and the known standard deviation Unknown standard deviation plans are based upon the sample mean and the sample standard deviation. Average range plans are based upon the sample mean and the average range in subsamples

This paper presents a matching collection of variables sampling plans based on known standard deviation, unknown standard deviation, and average range Each operating characteristic curve shown represents a variables sampling plan of any one of the three types. In other words, if the user chooses an OC curve he has at his disposal the choice of the three types of plans, all guaranteeing essentially the same protection.

* This work was performed under the sponsorship of the office of Naval Research The authors are indebted to Albert H. Bowker for invaluable aid.

For the three types of plans to have approximately the same OC curve it is necessary in some cases to use a different sample size for each type. Therefore, it is not possible to index the plans by sample size. Instead the plans are indexed by code letter and Acceptable Quality Level (AQL). The AQL may be defined as that quality level considered acceptable as a process average. Each combination of code letter and AQL refers to a single OC curve which is used as the OC curve for any one of the three types of plan The probability of acceptance at the AQL varies from 89 for the code letter representing the smallest sample sizes to 99 for the code letter representing the largest sample sizes This variation in the probability of acceptance at the AQL follows the practice of Military Standard 105 A[1] [9].

Like attribute plans these variables plans involve measurement of lot quality by the per cent defective Previous tables of variables plans involve measurement of lot quality by the average and variability of the measurements. With the plans given here it is not necessary to shift to attributes sampling whenever per cent defective is the appropriate measure.

All variables sampling procedures for two-sided specification limits suffer from the fact that the probability of accepting a submitted lot with given per cent defective p does not depend on p alone but on the division of p into two components, the per cent lying above the upper specification limit and the per cent lying below the lower specification limit For this reason a two-sided procedure does not have a unique OC curve but rather a band of curves, each curve within the band representing a possible division of p With the two-sided procedure[2] given in this paper this band is so narrow as to be for all practical purposes a single curve. Since the OC curve for the one-sided test is contained within this narrow band, it is used as the OC curve for the two-sided procedure.

II GENERAL INSPECTION CRITERIA

Associated with each inspection characteristic are the design specifications. If only an upper specification limit U is given, the item is considered defective if its measurement exceeds U. If only a lower specification L is given, the item is considered defective if its measurement is smaller than L. If both upper and lower limits are specified, the item is considered defective if its measurement either exceeds U or is smaller than L.

---

[1] Military Standard 105 A, usually referred to as MIL-STD 105 A, is a U S Department of Defense document containing acceptance sampling plans by attributes.

[2] The graphical two-sided procedure for unknown standard deviation plans given by Bowker and Goode [2] is equivalent to the two-sided procedure for unknown standard deviation plans presented in this paper.

If the per cent defective of a submitted lot is known, no sampling inspection is necessary to determine whether or not the lot is to be accepted. If the per cent defective is sufficiently small, the lot is accepted, otherwise it is rejected Since such knowledge about the per cent defective is rare, a logical procedure is to *estimate* the per cent defective from a sample, and accept or reject the lot on the basis of this estimate A sampling plan is then described as consisting of the sample size n; a method of estimating the per cent defective, and a "maximum allowable estimated per cent defective," p^*. If only an upper specification limit, U, is given the estimate of the percentage above this limit, $\hat{p}_U$, is obtained from the sample of size n. If $\hat{p}_U \leqq p^*$, the lot is accepted. If only a lower specification limit, L, is given, the estimate of the percentage below this limit, $\hat{p}_L$, is obtained from the sample of size n. If $\hat{p}_L \leqq p^*$, the lot is accepted. If a double specification limit is given, both $\hat{p}_U$ and $\hat{p}_L$ are computed. If $\hat{p} = \hat{p}_U + \hat{p}_L \leqq p^*$, the lot is accepted.

III KNOWN STANDARD DEVIATION PLANS

In this section, it will be assumed that the standard deviation σ of the measurements is known. A complete set of sampling plans indexed according to AQL and code letter[3] is given in Table 1. Operating characteristic curves for these plans are the graphs of Figures 1–16.

Before presenting the general inspection procedure for known standard deviation plans, some discussion about two-sided specifications is in order For this case information is available about the per cent defective in the lot before a sample is drawn There is a sufficiently small value of $(U-L)/\sigma$, which is a function only of the AQL, that assures incoming quality worse than the AQL Consequently, before sampling $(U-L)/\sigma$ should be calculated and the lot rejected (with no sample taken) if this value is smaller than the minimum value of $(U-L)/\sigma$, on the grounds that the incoming quality is poorer than the AQL. Minimum values of $(U-L)/\sigma$ can be found in Table 1 Intuitively, this implies that the combination of narrow specification limits and large standard deviation mitigate against the submittal of good quality. The inspection procedure is as follows:

1 Draw a random sample of n items and compute

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

[3] The code letters run from B to Q, omitting A. This was done so that these tables would resemble the tables in MIL-STD 105 A.

2. a. For an upper specification limit compute

$$C_U' = \left(\frac{U - \bar{x}}{\sigma}\right)v \quad \text{where} \quad v = \sqrt{\frac{n}{n-1}}\,.$$

b. For a lower specification limit compute

$$C_L' = \left(\frac{\bar{x} - L}{\sigma}\right)v \quad \text{where} \quad v = \sqrt{\frac{n}{n-1}}\,.$$

c. For a double specification limit, compute both

$$C_U' = \left(\frac{U - \bar{x}}{\sigma}\right)v \quad \text{and} \quad C_L' = \left(\frac{\bar{x} - L}{\sigma}\right)v$$

if $(U-L)/\sigma$ is greater than the minimum allowable value found in Table I Otherwise, the lot is rejected before a sample is drawn.

3 Enter Table II with C_U' and/or C_L' and read out $\hat{p}_U$ and/or $\hat{p}_L$ whichever is applicable [4]

4. For an upper specification limit accept the lot if $\hat{p}_U \leqq p^*$. For a lower specification limit accept the lot if $\hat{p}_L \leqq p^*$. For a double specification limit accept the lot if $\hat{p}_U + \hat{p}_L \leqq p^*$.

Example. The specified minimum yield point for certain steel castings is 55,000 pounds per square inch The standard deviation is known to be $\sigma = 3{,}000$ psi. A 1% AQL plan is to be used with a sample size of 6. The yield points of the sample specimens are

62,000; 61,000; 68,500; 59,500; 65,500; 63,900.

The following are computed from the data:

1 $\bar{x} = 63{,}400$

2b $C_L' = (63{,}400 - 55{,}000/3{,}000)(1\ 095) = 3\ 07$.

3 From Table II $\hat{p}_L = .107\%$.

4. From Table I p^* is 2 57% Hence the lot is accepted.

IV. UNKNOWN STANDARD DEVIATION PLANS

In this section it will be assumed that the standard deviation of the measurement is unknown A complete set of sampling plans indexed according to AQL and code letter is given in Table III. Operating characteristics curves for these plans are provided in the graphs of Figures 1–16

[4] It is shown in Section VI-3 that these estimates are the uniformly minimum variance unbiased estimates of the true per cent defective.

The inspection procedure is as follows:

1. Draw a random sample of n items and compute

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ and } s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}.$$

2. a. For an upper specification limit, compute $C_U = (U - \bar{x})/s$.
   b. For a lower specification limit, compute $C_L = (\bar{x} - L)/s$.
   c. For a double specification limit, compute both $C_U = (U - \bar{x})/s$ and $C_L = (\bar{x} - L)/s$
3. Enter Table IV with C_U and/or C_L and read out $\hat{p}_U$ and/or $\hat{p}_L$ whichever is applicable.[5]
4. For an upper specification limit accept the lot if $\hat{p}_U \leqq p^*$. For a lower specification limit accept the lot if $\hat{p}_L \leqq p^*$. For a double specification limit accept the lot if $\hat{p}_U + \hat{p}_L \leqq p^*$.

Example The specifications for electrical resistance of a certain electrical component is $U = 660$ ohms and $L = 620$ ohms. A 4% AQL plan is to be used with a sample of size 10 The data are as follows:

639, 640, 650, 647, 662, 637, 652, 643, 657, 649

The following are computed from the data.

1. $\bar{x} = 647.6$, $s = \sqrt{65.38} = 8.04$

2c $C_L = 647.6 - 620/8.04 = 3.44$, $C_U = 660 - 647.6/8.04 = 1.54$.

3 From Table IV

$$\hat{p}_L = 0,$$
$$\hat{p}_u = 5.31\%,$$
$$\hat{p} = \hat{p}_L + \hat{p}_u = 5.31\%$$

4 From Table III

$$p^* = 1.30\%.$$

Since $\hat{p} > p^*$, the lot is rejected For possible review board evidence, the best estimate of the per cent defective in the lot is 5.31%. It is evident that the total estimated per cent defective is due almost entirely to items exceeding the upper specification limit.

V. AVERAGE RANGE PLANS

In this section it will be assumed that the standard deviation of the measurement is unknown. A complete set of sampling plans indexed

[5] It is shown in Section VI-4 that these estimates are the uniformly minimum variance unbiased estimates of the true per cent defective.

according to AQL and code letter is given in Table V. Operating characteristic curves for these plans are provided in the graphs of Figures 1–16.

The inspection procedure is as follows:

1. Draw a random sample of n items and group the measurements into k subgroups of 5 (where $n=5k$) with the exception of sample sizes of 3, 4, and 7 For these sample sizes use a single subgroup of 3, 4, and 7, respectively. For each subgroup compute the range (R_i). Determine the average range

$$\bar{R} = \left(\sum_{i=1}^{k} R_i\right) \Big/ k.$$

Calculate

$$\bar{x} = \left(\sum_{i=1}^{n} x_i\right) \Big/ n$$

2. a For an upper specification limit, compute

$$C_U'' = \left(\frac{U - \bar{x}}{\bar{R}}\right) h\,^{6}$$

b. For a lower specification limit, compute

$$C_L'' = \left(\frac{\bar{x} - L}{\bar{R}}\right) h\,^{6}$$

c. For a double specification limit, compute both

$$C_U'' = \left(\frac{U - \bar{x}}{\bar{R}}\right) h \quad \text{and} \quad C_L'' = \left(\frac{\bar{x} - L}{\bar{R}}\right) h.^{6}$$

3. Enter Table VI with C_U'' and/or C_L'' and read out $\hat{p}_U$ and/or $\hat{p}_L$ whichever is applicable
4. For an upper specification limit accept the lot if $\hat{p}_U \leqq p^*$ For a lower specification limit accept the lot if $\hat{p}_L \leqq p^*$. For a double specification limit accept the lot if $\hat{p}_U + \hat{p}_L \leqq p^*$.

Example. The maximum temperature of operation for a certain device is specified as 180°. A 4% AQL plan is to be used with a sample of size 15. The sample items have operative temperatures of

$$178,\ 175,\ 174,\ 158,\ 172, \qquad R_1 = 20,$$
$$177,\ 166,\ 172,\ 167,\ 163, \qquad R_2 = 14,$$
$$174,\ 173,\ 162,\ 182,\ 170, \qquad R_3 = 20.$$

[6] h is a factor found in Table V which is a function of the sample code letter.

The following are computed from the data:

1. $\bar{x} = 170.87$.

2b. $C_L'' = (180 - 170\ 87/18)2.394 = 1\ 21$.

3. $\hat{p}_L = 11\ 08\%$.

4. From Table VI

$$p^* = 9\ 61\%.$$

Since $\hat{p}_L > p^*$ the lot is rejected.

VI-1 GENERAL THEORY[7]

Blackwell [1] has shown that if x has density $p\theta(x)$, g is any unbiased estimate of θ, and T is a sufficient statistic for θ, then $E(g|T)$ is also unbiased, and furthermore, has a variance no greater than that of g. Lehmann and Scheffé [8] have extended this result and have shown that if T is complete,[8] then every estimable function[9] $h(\theta)$ possesses an unbiased estimate with uniformly smallest variance and this estimate is the unique unbiased estimate of $h(\theta)$ which is a function of T. Heuristically, a statistic T is complete means that if there is a function of T which has expected value equal to zero, then this function must be identically zero for all values of T for which the function has positive probability

It has already been shown [8] that the sufficient statistics for the normal distribution are complete Furthermore, it is evident that there exists an unbiased estimate of the fraction of a normal population lying outside a fixed interval, e g , the fraction of independent observations in a sample from this normal population which lie outside this interval Consequently, the fraction of a normal population lying outside a fixed interval (p) is an estimable function and has an unbiased estimate with uniformly smallest variance, and this estimate is the unique unbiased estimate of p which is a function of the sufficient statistics

We are concerned with estimating the parameter

$$p = 1 - \int_L^U \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)((z-\mu)^2/\sigma^2)} dz$$

Two cases will be considered, namely, (1) μ unknown, σ known (2) both μ and σ unknown

[7] The idea of using the uniformly minimum variance estimate of the per cent defective as the test statistic for two-sided variables acceptance procedures is due to Albert H Bowker, Lincoln Moses, and Herman Rubin

[8] A statistic T (more properly a family of distributions of T) is said to be complete if $E_\theta[f(T)] = 0$ for all $\theta\epsilon r$ implies $f(T) = 0$ except possibly on a set N for which $p_\theta(N) = 0$ for all $\theta\epsilon r$

[9] Given a random variable x with density $p_\theta(x)$, a function $h(\theta)$ is said to be estimable if there exists a function $g(x)$ such that $E_\theta[g(x)] = h(\theta)$

Let $x_1, x_2, \cdots, x_n$ be independent random variables from the above normal distribution. Define $\tilde{p}'$ as the usual attribute estimate of the fraction defective i e., the ratio of the number of defective items to the sample size. It is evident that $\tilde{p}'$ is unbiased Hence it follows from Blackwell [1] and Lehmann and Scheffé's [8] results that $\hat{p}=E(\tilde{p}'\mid T)$ is the unique uniformly minimum variance (UMV) unbiased estimate of p, where T are the sufficient statistics for the normal distribution. Since $\tilde{p}'$ is the sum of independent identically distributed random variables taking on the values 0 and 1, it is evident that $\hat{p}=E(\tilde{p}'\mid T)$ is equivalent to $E(\tilde{p}\mid T)$ where $\tilde{p}$ is defined as follows. Let y be any one of the observations $(x_1, x_2, \cdots, x_n)$ say x_1.

$$\tilde{p}(y, x_2, x_3, \cdots, x_n) = 0, \tag{1}$$

if $L \leqq y \leqq U$

$$\tilde{p}(y, x_2, x_3, \cdots, x_n) = 1,$$

otherwise.

VI-2. THE ESTIMATE WHEN THE POPULATION VARIANCE IS KNOWN

If the observations are drawn from a normal population with unknown mean μ, and known variance σ^2, then

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is a sufficient statistic In this case $\hat{p}=E(\tilde{p}\mid\bar{x})$ is the UMV unbiased estimate of p If $\tilde{p}$ is the estimate (1), then

$$\begin{aligned}\hat{p}(\bar{x}) &= E(\tilde{p}\mid\bar{x}) = \Pr\{\tilde{p} = 1 \mid \bar{x}\} = 1 - \Pr\{L \leqq y \leqq U \mid \bar{x}\} \\ &= 1 - \int_L^U \frac{g(y \mid \bar{x})}{h(\bar{x})}\,dy\end{aligned}$$

where $g(y, \bar{x})$ is the joint probability density of y and $\bar{x}$, and $h(\bar{x})$ is the probability density of $\bar{x}$

Consider the joint probability density of y and

$$\bar{x}' = \sum_{i=2}^{n} \frac{x_i}{n-1}$$

$$\frac{\sqrt{n-1}}{2\pi\sigma^2}\, e^{-1/(2\sigma^2)[(y-\mu)^2+(n-1)(\bar{x}'-\mu)^2]}$$

The transformation

$$\bar{x} = \frac{(n-1)\bar{x}'}{n} + \frac{y}{n}$$

leads to

$$g(y, \bar{x}) = \frac{n}{2\pi\sigma\sqrt{n-1}}\, e^{-n/(2\sigma^2)\{(\bar{x}-\mu)^2+(y-\bar{x})^2/\}}$$

and division by

$$h(\bar{x}) = \frac{\sqrt{n}}{\sqrt{2\pi}\sigma}\, e^{-n(\bar{x}-\mu)^2/2\sigma^2}$$

results in

$$\frac{g(y, x)}{h(\bar{x})} = \sqrt{\frac{n}{n-1}}\, \frac{1}{\sqrt{2\pi}\sigma}\, e^{-n(y-\bar{x})^2/2\sigma^2(n-1)}.$$

Therefore

$$\begin{aligned}\hat{p}(\bar{x}) &= 1 - \int_L^U \sqrt{\frac{n}{n-1}}\, \frac{1}{\sqrt{2\pi}\sigma}\, e^{-n(y-\bar{x})^2/2\sigma^2(n-1)} dy \\ &= \int_{-\infty}^{\sqrt{n/(n-1)}(L-\bar{x})/\sigma} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt \\ &\quad + \int_{\sqrt{n/(n-1)}(U-\bar{x})/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}}\, e^{-t^2/2} dt.\end{aligned}$$

VI-3. THE KNOWN-σ ACCEPTANCE CRITERION

The acceptance procedure is formulated as follows. Accept a lot of items if in the sample $\hat{p} \leqq p^*$ where p^* is so chosen that if the population percentage defective is p, i e , if the portion of the population lying outside (U, L) is p, then the probability of acceptance will be L_p. It is shown that in the one-sided case this is equivalent to the well-known and widely used procedure: Accept if $\bar{x} \leqq U - k\sigma$.

Let K_ϵ be defined by

$$\int_{K_\epsilon}^{\infty} \frac{e^{-t^2/2}}{\sqrt{2\pi}} = \epsilon.$$

In the one-sided case, when $L = -\infty$, then

$$\hat{p} = \int_{-\sqrt{n/(n-1)}\,(U-\bar{x})/\sigma}^{\infty} \leqq p^*$$

implies that

$$\sqrt{\frac{n}{n-1}}\,\frac{U-\bar{x}}{\sigma} \geqq K_{p^*} \quad \text{or} \quad \bar{x} \leqq U - \sqrt{\frac{n-1}{n}}\,K_{p^*}\sigma.$$

If we take

$$k = \sqrt{\frac{n-1}{n}}\,K_{p^*},$$

the OC curves of the two acceptance procedures will be identical

For the case where there is double specification limit and the standard deviation is known, it is evident that if

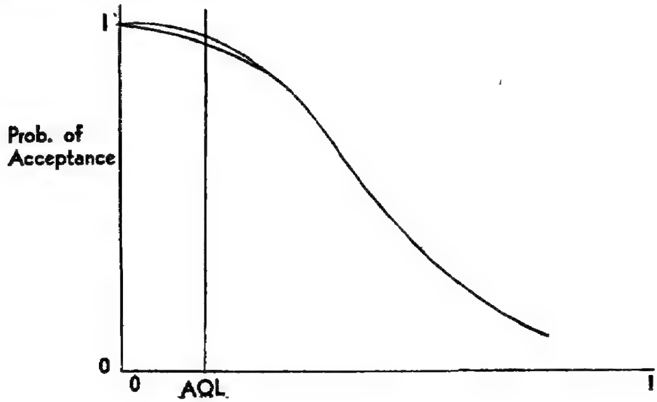
$$\frac{U-L}{\sigma} < 2K_{AQL/2},$$

the incoming value of p, the per cent defective, is greater than the AQL. Consequently, the lot should be rejected without a sample being drawn The OC curve of this two sided test takes the form of a band, the lower bound of which is the one sided OC curve For incoming quality which is no worse than the AQL the upper bound corresponds to equal division of the per cent defective The band here is relatively wide, but this is desirable since this gives "good" quality a better chance of acceptance For quality worse than the AQL, the above restriction on $(U-L)/\sigma$ eliminates certain divisions, including equal division, so that the band quickly tends to the one sided curve A diagram of the OC band is shown at the top of the next page.

VI-4 THE ESTIMATE OF p WHEN THE POPULATION VARIANCE IS UNKNOWN

If the observations are drawn from a normal population with unknown mean μ and unknown variance σ^2, then the pair of sample values $\bar{x}$, and $\sum (x_i - \bar{x})^2$ are sufficient statistics In this case $\hat{p}(\bar{x}, S^2) = E(\tilde{p} \mid \bar{x}, S^2)$ where

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}, \qquad S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2.$$



Per cent defective.

If $\hat{p}$ is the estimate (1), then

$$\hat{p}(\bar{x}, S^2) = \Pr\left\{\hat{p} = 1 \mid \bar{x}, S^2\right\} = 1 - \int_L^U \frac{f(y, \bar{x}, S^2)}{h(\bar{x}, S^2)}\, dy$$

where $f(y, \bar{x}, S^2)$ is the joint probability density of y, one of the observations in the sample, the sample mean $\bar{x}$, and the sample sum of squares S^2, and $h(\bar{x}, S^2)$ is the joint probability density of $\bar{x}$ and S^2.

It is well known that $h(\bar{x}, S^2)$ is given by

$$h(\bar{x}, S^2) = \frac{\sqrt{n}(S^2)^{(n-3)/2} e^{-n(\bar{x}-\mu)^2/2\sigma^2 - S^2/(2\sigma^2)}}{\sqrt{2\pi}\sigma\Gamma\left(\frac{n-1}{2}\right)(2\sigma^2)^{(n-1)/2}}.$$

To find $f(y, \bar{x}, S^2)$ consider the joint density of the sample This may be expressed as the joint density of the mutually independent sample statistics y, $\bar{x}'$, and S'^2 where

$$\bar{x}' = \sum_{i=2}^{n} \frac{x_i}{n-1}, \qquad S'^2 = \sum_{i=2}^{n} (x_i - \bar{x}')^2,$$

$$\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}\right) \cdot \left(\frac{\sqrt{n-1}}{\sqrt{2\pi}\sigma} e^{-(n-1)(\bar{x}'-\mu)^2/(2\sigma^2)}\right)$$

$$\left[\frac{(S'^2)^{(n-4)/2}}{\Gamma\left(\frac{n-2}{2}\right)(2\sigma^2)^{(n-2)/2}} e^{-S'^2/2\sigma^2} \right]. \tag{2}$$

Now, let

$$\bar{x}' = \frac{n\bar{x}}{n-1} - \frac{y}{n-1},$$

$$S'^2 = S^2 - \frac{n}{n-1}(\bar{x} - y)^2,$$

$$y = y.$$

Under this transformation the expression (2) becomes

$$K\left\{S^2 - \frac{n}{n-1}(\bar{x} - y)^2\right\}^{(n-4)/2} e^{-(n-1)/2\sigma^2\{(y-\mu)^2/(n-1)+((n\bar{x}-y)/(n-1)-\mu)^2+S^2/(n-1)-n(\bar{x}-y)^2/(n-1)^2\}} \tag{3}$$

where

$$K = \frac{n}{\sqrt{n-1}\; 2\pi\sigma^2(2\sigma^2)^{(n-2)/2}\Gamma\left(\frac{n-2}{2}\right)},$$

and the variables are subject to the restrictions

$$-\infty < y < \infty$$

$$0 \leqq S^2 \leqq \infty$$

$$\left|\frac{\bar{x} - y}{S}\right| \leqq \sqrt{\frac{n-1}{n}}.$$

Dividing the expression (3) by $h(\bar{x}, S^2)$ results in the conditional density of y, given $\bar{x}$ and S^2.

$$\sqrt{\frac{n}{n-1}}\; \frac{\Gamma\left(\frac{n-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{n-2}{2}\right)}\; \frac{1}{S}\left\{1 - \left(\frac{\bar{x} - y}{S}\right)^2 \frac{n}{n-1}\right\}^{(n-4)/2}. \tag{4}$$

If now the transformation

$$z = \frac{1}{2} + \frac{1}{2}\,\frac{\bar{x} - y}{S}\sqrt{\frac{n}{n-1}}$$

is made, the density of the random variable z is obtained:

$$(5)\qquad g(z) = \frac{\Gamma(n-2)}{\Gamma\left(\frac{n-2}{2}\right)\Gamma\left(\frac{n-2}{2}\right)} z^{(n/2-1)-1}(1-z)^{(n/2-1)-1}$$

$$0 \leqq z \leqq 1.$$

This will be recognized as a symmetrical Beta distribution with parameters $(n/2)-1$. Hereafter, $g(z)dz$ will be denoted by

$$d\beta\left(\frac{n}{2}-1\right).$$

Since $\hat{p}(\bar{x}, S^2) = \Pr\{y > U \mid \bar{x}, S^2) + \Pr\{y < L \mid \bar{x}, S^2)$, then in terms of z

$$\begin{aligned}\hat{p}(\bar{x}, S^2) &= \Pr\left\{z \leqq \frac{1}{2} - \frac{1}{2}\,\frac{U-\bar{x}}{S}\sqrt{\frac{n}{n-1}}\right\} \\ &\quad + \Pr\left\{z \geqq \frac{1}{2} - \frac{1}{2}\,\frac{x-L}{S}\sqrt{\frac{n}{n-1}}\right\} \\ &= \int_0^{\max[0,1/2-(U-\bar{x})\sqrt{n}/2S\sqrt{n-1}]} d\beta\left(\frac{n}{2}-1\right) \\ &\quad + \int_0^{\max[0,1/2-(\bar{x}-L)\sqrt{n}/2S\sqrt{n-1}]} d\beta\left(\frac{n}{2}-1\right).\end{aligned}$$

VI-5. THE UNKNOWN-σ ACCEPTANCE CRITERION

Just as in the case where the population variance is known the acceptance procedure is: Accept if $\hat{p}(\bar{x}, S^2) \leqq p^*$ where p^* is so chosen that if the population fraction defective is p, the probability of acceptance will be L_p. It is shown that in the one-sided case, i e , where there is only one specification limit, say U, this procedure is equivalent to the well-known and widely used test $\bar{x}+ks \leqq U$, where $s = S/\sqrt{n-1}$

Let $\beta_p{}^*$ be defined by $\int_0^{\beta_{p^*}} d\beta(n/2-1) = p^*$, then $\hat{p} \leqq p^*$ if and only if

$$\frac{1}{2} - \frac{1}{2}\sqrt{\frac{n}{n-1}}\,\frac{U-\bar{x}}{S} \leqq \beta_{p^*}$$

and since $s = S/\sqrt{n-1}$ this is equivalent to

$$\bar{x} + \frac{n-1}{\sqrt{n}}(1-2\beta_{p^*})s \leq U.$$

If k is taken equal to $(n-1)/\sqrt{n}(1-2\beta_{p^*})$, the procedures $\hat{p} \leqq p^*$ and $\bar{x}+ks \leqq U$ will have identical OC curves. In this paper the acceptance criteria $\hat{p}$ based on the statistic $(U-\bar{x})/s$ are tabled rather than $(U-\bar{x})/S$, to conform with the common use of the square root of the unbiased estimator of σ^2.

For the case where there is a double specification limit one of the authors made numerical investigations [11] of the OC curves for various divisions of the per cent defective It was evident that the band was so narrow that for all practical purposes it can be assumed to be a single OC curve, i e , the OC curve of the one sided plan.

VI-6 THE ACCEPTANCE PROCEDURE AND ESTIMATE BASED ON THE RANGE x

Since the pair of sample statistics $(\bar{x}, \overline{R})$, where $\overline{R}$ is the sample range or average range,[10] is not sufficient for the normal distribution when the mean and variance are both unknown, no uniformly minimum variance estimate of p which is a function of $\bar{x}$ and $\overline{R}$ can be derived. In this paper the following estimate, $\hat{p}(\bar{x}, \overline{R})$, is used

$$\hat{p}(\bar{x}, \overline{R}) = \int_0^{\max\,[0,1/2-\alpha(U-\bar{x})/\nu\overline{R}]} d\beta\left[\frac{\nu+1}{2}-1\right] + \int_0^{\max\,[0,1/2-\alpha(\bar{x}-L)/\nu\overline{R}]} d\beta\left[\frac{\nu+1}{2}-1\right]$$

where α and ν are constants for fixed n, which will be explained in a subsequent paragraph of this section

No difficulty is encountered in connection with OC curves (which were obtained by numerical integration) for the acceptance procedure $\hat{p}(\bar{x}, \overline{R}) \leqq p^*$ since $\Pr\{\hat{p} \leqq p^*\} = \Pr\{\bar{x}+k\overline{R} \leqq U\}$ whenever $k=\nu(1-2\beta_{p^*})/\alpha$. The procedure to accept if $\bar{x}+k\overline{R} \leqq U$ is of course the one commonly used in place of the sample standard deviation in variables sampling

A heuristic justification for the statistic $\hat{p}(\bar{x}, \overline{R})$ as an estimate of p is the following. It has been verified by numerical investigation that the statistic $\alpha(U-\bar{x})/\overline{R}$ is approximately distributed as non-central t with degrees of freedom ν and eccentricity $\sqrt{\nu+1}(U-\mu)/\sigma$, whenever μ and σ are such that $(U-\mu)/\sigma$ is sufficiently large, i e , when the population fraction defective is small. The numbers α and ν are con-

[10] Whenever the total sample size is a multiple of 5, the subgroup size is taken as 5, so that $\overline{R}$ is the average range of m subgroups of S (the sample consisting of Sm observations) For sample sizes of 3, 4, and 7, $\overline{R}$ is taken as the range of the sample

stants for each fixed sample size n and are determined as follows. Under the assumption that $\alpha(U-\bar{x})/\overline{R}$ is distributed as non-central $t_{\nu, \sqrt{\nu+1}K_p}$ for small p then

$$\Pr\left\{\bar{x} + \frac{k}{\alpha}\overline{R} \leqq U\right\} = \Pr\left\{x_{\nu+1} + \frac{k}{\sqrt{\nu+1}} s_\nu \leqq U\right\}$$

where $\bar{x}_{\nu+1}$ and s_ν are the sample mean and sample standard deviation based on ν observations Equating the first two moments of the statistics $\bar{x}+(k/\alpha)\overline{R}$ and $\bar{x}_{\nu+1}+(k/\sqrt{\nu+1})s_\nu$ and solving for α and ν gives α as a function of ν and ν as a function of n and k. In fact $\alpha=(d_2\sqrt{v})/c_2$ where d_2 is defined by $E\overline{R}=d_2\sigma$ and c_2 is defined by $E(s)=c_2\sigma$ In order to make the constants independent of k the values for the limiting case, i e., when p goes to zero, were chosen [11]

This approximation was verified by the authors by numerical integration for the statistic $(U-\bar{x})/\overline{R}$ and reference to tables of the non-central t It was found to be very good for both large and small n for most of the AQL values used in this report It is especially good for small n regardless of p For the larger values of p where n is large the approximation did not hold with great accuracy However, it could have been improved by making ν, and hence α, functions of p as well as of n This was not done because this would have entailed increasing the number of tables of estimates, thereby complicating use of the tables by sheer bulk. Furthermore, the effect of this on the estimate $\hat{p}(\bar{x}, \overline{R})$ would have been slight since on the average the sample values of $(U-\bar{x})/\overline{R}$ fall in the region of the tables where the estimate $\hat{p}$ changes very little as n changes.

Essentially the above argument indicates that $\hat{p}(\bar{x}, \overline{R})$ is approximately distributed as $\hat{p}(\bar{x}_{\nu+1}, s_\nu)$ Hence $\hat{p}(\bar{x}, \overline{R})$ has sampling properties similar to that of $\hat{p}(\bar{x}_\nu, s_\nu)$ where the effective number of observations is $\nu+1$.

VI-7 METHODS USED IN COMPUTING THE OPERATING-CHARACTERISTIC CURVES

The OC curves shown in Figures 1–16 were computed for one-sided sampling procedures based on the estimate $\hat{p}(\bar{x}, s)$ of Section VI-4. As mentioned previously, $\hat{p} \leqq p^*$ if and only if $\bar{x}+ks \leqq U$ where k is given

[11] This value of ν obtained in this way is exactly that obtained by Patnaik who suggested that $\sqrt{n}c(U-\bar{x})/\overline{R}$ may be approximated by non-central t with degrees of freedom ν but with eccentricity $\sqrt{n}(U-\mu)/\sigma$ This approximation is excellent and will hold for all p but would not fit into the framework of the proposed estimate $\hat{p}(\bar{x}, \overline{R})$ See [10].

by the relation $k=(n-1)(1-2\beta_{p^*})/\sqrt{n}$ and β_{p^*} is defined by

$$\int_0^{\beta_{p^*}} d\beta\left(\frac{n}{2}-1\right) = p^*.$$

Therefore

$$\Pr\{\hat{p} \leqq p^*\} = \Pr\{\bar{x}+ks \leqq U\} = \Pr\left\{\frac{U-\bar{x}}{s} \geqq k\right\}$$

$$= \Pr\left\{\frac{\dfrac{\sqrt{n}(U-\mu)}{\sigma} - \dfrac{\sqrt{n}(\bar{x}-\mu)}{\sigma}}{s/\sigma} \geqq \sqrt{n}k\right\}.$$

The quantity on the left inside the last bracket has the non-central t distribution with degrees of freedom $f=n-1$ and eccentricity $\delta=\sqrt{n}(U-\mu)/\sigma$. If it is given that the fraction of a normal population exceeding U is equal to p, then $(U-\mu)/\sigma=K_p$ where K_p is defined by

$$\int_{K_p}^{\infty} \frac{e^{-t^2/2}}{\sqrt{2\pi}} = p$$

In this case the parameter δ is equal to $\sqrt{n}K_p$. The points on the OC curves were computed by finding in the Johnson and Welch tables [7] the probability that a non-central t variate with parameters $n-1$ and $\sqrt{n}K_p$ would exceed the value $\sqrt{n}k$.

To include separate sets of OC curves for the average-range plans and the known-σ plans would have been impracticable simply because of sheer bulk. Therefore it was deemed necessary to use the same sets of graphs for all three types of sampling plans as well as for the two-sided plans.

In the case of known-σ this was done by altering the sample sizes and the values of the acceptance criteria so as to obtain OC curves passing through two selected points on the curves given in Figures 1–16 It is evident that specifying two points on the OC curve of a sampling plan uniquely determines the sample size n and the value of the criterion p^*. Since for a known-σ plan $\hat{p}(\bar{x})\leqq p^*$ if and only if $\bar{x}\leqq U-\sqrt{(n-1)/n}K_{p^*}\sigma$, and $\bar{x}$ is a normally distributed variate, it was only necessary to solve the equations

$$\sqrt{n}K_{p_1} - \sqrt{n-1}K_{p^*} = L_{p_1}$$
$$\sqrt{n}K_{p_2} - \sqrt{n-1}K_{p^*} = L_{p}$$

for n and p^* in order to obtain an OC curve passing through the two points (p_1, L_{p_1}) and (p_2, L_{p_2}). In particular, the curves were matched for the point for which $p_1 = \text{AQL}$, and the point for which $L_{p_2} = .10$. Since the sample size is necessarily an integer and the solutions of the equations do not yield integer values for n, the match is not exact. However for $n \geqq 10$ it is exact within the limits of accuracy inherent in reading the graphs.

In the case of average-range plans the above procedure was altered in the following ways. For sample size code letters B through G the same sample size was used as in the corresponding plans based on the sample standard deviation The acceptance criterion was changed so that the probability of acceptance for a range plan was the same at the AQL as for a standard deviation plan Then a few additional points on the resulting OC curve were computed to check on the magnitude of the deviation between the two curves. For no case was this deviation so large that the resulting curve was closer to the adjacent curves on the graph than it was to the one it was intended to match For sample size code letters beyond G a shift was made to a new sample size for each code letter in order to facilitate the matching which was then carried out as in the case of code letters B through G All computations used in this connection involved computing $\Pr\ \{\bar{x}+k\overline{R} \leqq K_p | p\}$ by numerical integration using tables of the probability distribution of the average-range which were obtained from [12].

(Tables and Charts are given on following pages)

MASTER TABLE FOR PLANS BASED ON KNOWN STANDARD DEVIATION

TABLE I

Code letter	Acceptable Quality Levels: .04			.065			.10			.15			.25			.40			.65		
Min. Value $\frac{U-L}{\sigma}$	7.08			6.82			6.58			6.35			6.05			5.76			5.44		
	n	p^*	v	n	p^*	v	n	p^*	v	n	p^*	v	n	p^*	v	n	p^*	v	n	p^*	v
B																					
C																					
D																			2	1.28	1.414
E													2	310	1.414	2	510	1.414	3	1.94	1.225
F										3	369	1.225	3	568	1.225	3	959	1.225	4	1.88	1.155
G	3	.079	1.225	3	.114	1.225	4	.290	1.155	4	.399	1.155	4	.681	1.155	5	1.09	1.118	5	1.76	1.118
H	4	.111	1.155	4	.161	1.155	5	.296	1.118	5	445	1.118	6	.721	1.095	6	1.14	1.095	7	1.75	1.080
I	5	.130	1.118	6	.230	1.095	6	.321	1.095	6	.478	1.095	7	.756	1.080	8	1.14	1.069	8	1.80	1.069
J	6	.145	1.095	6	.234	1.095	7	.343	1.080	7	.507	1.080	8	.791	1.069	9	1.18	1.061	10	1.79	1.054
K	7	.141	1.080	7	.226	1.080	8	.330	1.069	9	.469	1.061	9	.760	1.061	10	1.14	1.054	11	1.73	1.049
L	8	.153	1.069	8	.243	1.069	9	.351	1.061	10	.494	1.054	11	.768	1.049	12	1.15	1.045	13	1.76	1.041
M	10	.141	1.054	11	.217	1.049	11	326	1.049	12	.461	1.045	13	.721	1.041	14	1.08	1.038	16	1.62	1.033
N	14	.138	1.038	15	.211	1.035	16	.308	1.033	17	.438	1.031	19	.673	1.027	21	1.00	1.025	23	1.51	1.023
O	19	.134	1.027	20	.207	1.026	22	.296	1.024	23	.423	1.023	25	.655	1.021	27	.980	1.019	30	1.47	1.017
P	27	.129	1.019	30	.193	1.017	31	283	1.017	34	.397	1.015	37	.615	1.014	40	921	1.013	44	1.39	1.012
Q	37	.130	1.014	40	.196	1.013	42	.285	1.012	45	.402	1.011	49	.620	1.010	54	.920	1.009	59	1.39	1.009

Code letter	Acceptable Quality Levels: 1.00			1.50			2.50			4.00			6.50			10.00			15.00		
Min. Value $\frac{U-L}{\sigma}$	5.15			4.86			4.48			4.11			3.69			3.29			2.88		
	n	p^*	v	n	p^*	v	n	p^*	v	n	p^*	v	n	p^*	v	n	p^*	v	n	p^*	v
B																					
C	2	2.73	1.414	2	3.90	1.414	2	6.11	1.414	2	9.27	1.414	3	17.74	1.225	3	24.22	1.225	4	33.67	1.225
D	2	2.23	1.414	2	3.00	1.414	3	7.56	1.225	3	10.79	1.225	3	15.60	1.225	4	22.97	1.155	4	31.01	1.155
E	3	2.76	1.225	3	3.85	1.225	4	6.99	1.155	4	9.97	1.155	5	15.21	1.118	5	20.80	1.118	6	28.64	1.095
F	4	2.58	1.155	4	3.87	1.155	5	6.05	1.118	5	8.92	1.118	6	13.89	1.095	7	19.46	1.080	8	26.64	1.069
G	6	2.57	1.095	6	3.77	1.095	7	5.83	1.080	8	8.62	1.069	9	12.88	1.061	11	17.88	1.049	12	24.88	1.045
H	7	2.62	1.080	8	3.68	1.069	9	5.68	1.061	10	8.43	1.054	12	12.35	1.045	14	17.36	1.038	16	23.96	1.033
I	9	2.59	1.061	10	3.63	1.054	11	5.60	1.049	13	8.13	1.041	15	12.04	1.035	17	17.05	1.031	20	23.43	1.026
J	11	2.57	1.049	12	3.61	1.045	13	5.58	1.041	15	8.13	1.035	18	11.88	1.029	21	16.71	1.025	24	23.13	1.022
K	12	2.49	1.045	14	3.43	1.038	15	5.34	1.035	18	7.72	1.029	20	11.57	1.026	24	16.23	1.022	27	22.63	1.019
L	14	2.51	1.038	15	3.54	1.035	18	5.29	1.029	20	7.80	1.026	23	11.56	1.023	27	16.27	1.019	31	22.57	1.017
M	17	2.35	1.031	19	3.28	1.027	22	4.98	1.024	25	7.34	1.021	29	10.93	1.018	33	15.61	1.016	38	21.77	1.013
N	25	2.19	1.021	28	3.05	1.018	32	4.68	1.016	36	6.95	1.014	42	10.40	1.012	49	14.87	1.010	56	20.90	1.009
O	33	2.12	1.016	36	2.99	1.014	42	4.55	1.012	48	6.75	1.011	55	10.17	1.009	64	14.58	1.008	75	20.48	1.007
P	49	2.00	1.010	54	2.82	1.009	61	4.35	1.008	70	6.48	1.007	82	9.76	1.006	95	14.09	1.005	111	19.90	1.005
Q	65	2.00	1.008	71	2.82	1.007	81	4.34	1.006	93	6.46	1.005	109	9.73	1.005	127	14.02	1.004	147	19.84	1.003

TABLE II*

TABLE FOR ESTIMATING THE LOT PERCENTAGE DEFECTIVE FOR PLANS BASED ON KNOWN STANDARD DEVIATION

c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'		c_L' or c_U'	
.00	50.000																														
.01	49.601	.26	39.743	.51	30.503	.76	22.363	1.01	15.625	1.26	10.383	1.51	06.552	1.76	03.920	2.01	02.222	2.26	01.191	2.51	00.604	2.76	00.289	3.01	00.131	3.26	00.056	3.51	00.022	3.76	00.008
.02	49.202	.27	39.358	.52	30.153	.77	22.065	1.02	15.380	1.27	10.204	1.52	06.426	1.77	03.836	2.02	02.160	2.27	01.160	2.52	00.587	2.77	00.280	3.02	00.126	3.27	00.054	3.52	00.022	3.77	00.008
.03	48.803	.28	38.974	.53	29.806	.78	21.770	1.03	15.150	1.28	10.027	1.53	06.301	1.78	03.754	2.03	02.118	2.28	01.130	2.53	00.570	2.78	00.272	3.03	00.122	3.28	00.052	3.53	00.021	3.78	00.008
.04	48.405	.29	38.591	.54	29.460	.79	21.476	1.04	14.917	1.29	09.853	1.54	06.178	1.79	03.673	2.04	02.068	2.29	01.101	2.54	00.554	2.79	00.264	3.04	00.118	3.29	00.050	3.54	00.020	3.79	00.008
.05	48.006	.30	38.209	.55	29.116	.80	21.186	1.05	14.686	1.30	09.680	1.55	06.057	1.80	03.593	2.05	02.018	2.30	01.072	2.55	00.539	2.80	00.256	3.05	00.114	3.30	00.048	3.55	00.019	3.80	00.007
.06	47.608	.31	37.828	.56	28.774	.81	20.897	1.06	14.457	1.31	09.510	1.56	05.938	1.81	03.515	2.06	01.970	2.31	01.044	2.56	00.523	2.81	00.248	3.06	00.111	3.31	00.047	3.56	00.019	3.81	00.007
.07	47.210	.32	37.448	.57	28.434	.82	20.611	1.07	14.231	1.32	09.342	1.57	05.821	1.82	03.438	2.07	01.923	2.32	01.017	2.57	00.508	2.82	00.240	3.07	00.107	3.32	00.045	3.57	00.018	3.82	00.007
.08	46.812	.33	37.070	.58	28.096	.83	20.327	1.08	14.007	1.33	09.176	1.58	05.705	1.83	03.362	2.08	01.876	2.33	00.990	2.58	00.494	2.83	00.233	3.08	00.103	3.33	00.043	3.58	00.017	3.83	00.006
.09	46.414	.34	36.693	.59	27.760	.84	20.045	1.09	13.786	1.34	09.012	1.59	05.592	1.84	03.288	2.09	01.831	2.34	00.964	2.59	00.480	2.84	00.226	3.09	00.100	3.34	00.042	3.59	00.017	3.84	00.006
.10	46.017	.35	36.317	.60	27.425	.85	19.766	1.10	13.567	1.35	08.851	1.60	05.480	1.85	03.216	2.10	01.786	2.35	00.939	2.60	00.466	2.85	00.219	3.10	00.097	3.35	00.040	3.60	00.016	3.85	00.006
.11	45.620	.36	35.942	.61	27.093	.86	19.489	1.11	13.350	1.36	08.691	1.61	05.370	1.86	03.144	2.11	01.743	2.36	00.914	2.61	00.453	2.86	00.212	3.11	00.094	3.36	00.039	3.61	00.015	3.86	00.006
.12	45.224	.37	35.569	.62	26.763	.87	19.215	1.12	13.136	1.37	08.534	1.62	05.262	1.87	03.074	2.12	01.700	2.37	00.889	2.62	00.440	2.87	00.205	3.12	00.090	3.37	00.038	3.62	00.015	3.87	00.005
.13	44.828	.38	35.197	.63	26.435	.88	18.943	1.13	12.924	1.38	08.379	1.63	05.155	1.88	03.005	2.13	01.659	2.38	00.866	2.63	00.427	2.88	00.199	3.13	00.087	3.38	00.036	3.63	00.014	3.88	00.005
.14	44.433	.39	34.827	.64	26.109	.89	18.673	1.14	12.714	1.39	08.220	1.64	05.050	1.89	02.938	2.14	01.618	2.39	00.842	2.64	00.415	2.89	00.193	3.14	00.084	3.39	00.035	3.64	00.014	3.89	00.005
.15	44.038	.40	34.458	.65	25.785	.90	18.406	1.15	12.507	1.40	08.076	1.65	04.947	1.90	02.872	2.15	01.578	2.40	00.820	2.65	00.402	2.90	00.187	3.15	00.082	3.40	00.034	3.65	00.013	3.90	00.005
.16	43.644	.41	34.090	.66	25.463	.91	18.141	1.16	12.302	1.41	07.927	1.66	04.846	1.91	02.807	2.16	01.539	2.41	00.798	2.66	00.391	2.91	00.181	3.16	00.079	3.41	00.032	3.66	00.013	3.91	00.005
.17	43.251	.42	33.724	.67	25.143	.92	17.879	1.17	12.100	1.42	07.780	1.67	04.746	1.92	02.743	2.17	01.500	2.42	00.776	2.67	00.379	2.92	00.175	3.17	00.076	3.42	00.031	3.67	00.012	3.92	00.004
.18	42.858	.43	33.360	.68	24.825	.93	17.619	1.18	11.900	1.43	07.636	1.68	04.648	1.93	02.680	2.18	01.463	2.43	00.755	2.68	00.368	2.93	00.169	3.18	00.074	3.43	00.030	3.68	00.012	3.93	00.004
.19	42.465	.44	32.997	.69	24.510	.94	17.361	1.19	11.702	1.44	07.493	1.69	04.551	1.94	02.619	2.19	01.426	2.44	00.734	2.69	00.357	2.94	00.164	3.19	00.071	3.44	00.029	3.69	00.011	3.94	00.004
.20	42.074	.45	32.636	.70	24.196	.95	17.106	1.20	11.507	1.45	07.353	1.70	04.457	1.95	02.559	2.20	01.390	2.45	00.714	2.70	00.347	2.95	00.159	3.20	00.069	3.45	00.028	3.70	00.011	3.95	00.004
.21	41.683	.46	32.276	.71	23.885	.96	16.853	1.21	11.314	1.46	07.214	1.71	04.363	1.96	02.500	2.21	01.[illegible]	2.46	00.695	2.71	00.336	2.96	00.154	3.21	00.066	3.46	00.027	3.71	00.010	3.96	00.004
.22	41.294	.47	31.918	.72	23.576	.97	16.602	1.22	11.123	1.47	04.078	1.72	04.272	1.97	02.442	2.22	01.[illegible]	2.47	00.676	2.72	00.326	2.97	00.149	3.22	00.064	3.47	00.026	3.72	00.010	3.97	00.004
.23	40.905	.48	31.561	.73	23.270	.98	16.354	1.23	10.935	1.48	06.944	1.73	04.182	1.98	02.385	2.23	01.287	2.48	00.657	2.73	00.317	2.98	00.144	3.23	00.062	3.48	00.025	3.73	00.010	3.98	00.003
.24	40.517	.49	31.207	.74	22.965	.99	16.109	1.24	10.749	1.49	06.811	1.74	04.093	1.99	02.330	2.24	01.255	2.49	00.639	2.74	00.307	2.99	00.139	3.24	00.060	3.49	00.024	3.74	00.009	3.99	00.003
.25	40.129	.50	30.854	.75	22.663	1.00	15.866	1.25	10.565	1.50	06.681	1.75	04.006	2.00	02.275	2.25	01.222	2.50	00.621	2.75	00.298	3.00	00.135	3.25	00.058	3.50	00.023	3.75	00.009	4.00	00.003

* Values tabulated are read in per cent.

TABL. I

MASTER TABLE FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

Sample size code letter	Sample size	Acceptable Quality Levels													
		.04	.065	.10	.15	.25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00
		p*	p*	p*	p*	p*	p*	p*	p*	p*	p*	p*	p*	p*	p*
B	3										7.59	18.86	26.94	33.69	40.47
C	4								1.53	5.50	10.92	16.45	22.86	29.45	36.90
D	5							1.33	3.32	5.83	9.80	14.39	20.19	26.56	33.99
E	7					0.422	1.06	2.14	3.55	5.35	8.40	12.20	17.35	23.29	30.50
F	10				0.349	0.716	1.30	2.17	3.26	4.77	7.29	10.54	15.17	20.74	27.57
G	15	0.099	0.186	0.312	0.503	0.818	1.31	2.11	3.05	4.31	6.56	9.46	13.71	18.94	25.61
H	20	0.135	0.228	0.365	0.544	0.846	1.29	2.05	2.95	4.09	6.17	8.92	12.99	18.03	24.53
I	25	0.155	0.250	0.380	0.551	0.877	1.29	2.00	2.86	3.97	5.97	8.63	12.57	17.51	23.97
J	30	0.179	0.280	0.413	0.581	0.879	1.29	1.98	2.83	3.91	5.86	8.47	12.36	17.24	23.58
K	35	0.170	0.264	0.388	0.535	0.847	1.23	1.87	2.68	3.70	5.57	8.10	11.87	16.65	22.91
L	40	0.179	0.275	0.401	0.566	0.873	1.26	1.88	2.71	3.72	5.58	8.09	11.85	16.61	22.86
M	50	0.163	0.250	0.363	0.503	0.789	1.17	1.71	2.49	3.45	5.20	7.61	11.23	15.87	22.00
N	75	0.147	0.228	0.330	0.467	0.720	1.07	1.60	2.29	3.20	4.87	7.15	10.63	15.13	21.11
O	100	0.145	0.220	0.317	0.447	0.689	1.02	1.53	2.20	3.07	4.69	6.91	10.32	14.75	20.66
P	150	0.134	0.203	0.293	0.413	0.638	0.949	1.43	2.05	2.89	4.43	6.57	9.88	14.20	20.02
Q	200	0.135	0.204	0.294	0.414	0.637	0.945	1.42	2.04	2.87	4.40	6.53	9.81	14.12	19.92

All AQL and table values are in per cent defective

TABLE IV

TABLE FOR ESTIMATING THE LOT PERCENTAGE FOR PLANS BASED ON UNKNOWN STANDARD DEVIATION

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
0	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
.1	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.2	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.31	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.32	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.33	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.34	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.35	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.36	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.37	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.38	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.39	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.40	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.41	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.42	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.43	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.44	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.45	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.46	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.47	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.48	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.49	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.50	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.51	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.52	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.53	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.54	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.55	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.56	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.57	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.58	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.59	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.60	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.61	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.62	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.63	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.64	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.65	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.66	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.67	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.68	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.69	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
.70	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.71	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.72	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.73	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.74	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.75	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.76	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.77	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.78	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.79	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.80	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.81	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.82	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.83	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.84	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.85	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.86	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.87	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.88	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.89	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.90	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.91	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.92	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.93	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.94	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.95	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.96	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.97	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.98	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
.99	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.00	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.01	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.02	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.03	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.04	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.05	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.06	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.07	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.08	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.09	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

TABLE IV

TABLE FOR ESTIMATING THE LOT PERCENTAGE FOR PLANS BASED ON UNKNOWN STANDARD DEVIATION—(*Continued*)

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1.10	9.84	13.33	13.46	13.49	13.50	13.51	13.52	13.52	13.53	13.54	13.54	13.54	13.55	13.55	13.56	13.56
1.11	8.89	13.00	13.20	13.25	13.26	13.28	13.29	13.30	13.31	13.31	13.32	13.32	13.33	13.34	13.34	13.34
1.12	7.82	12.67	12.93	13.00	13.03	13.05	13.07	13.08	13.09	13.10	13.10	13.11	13.12	13.12	13.12	13.13
1.13	6.60	12.33	12.65	12.75	12.80	12.83	12.85	12.86	12.87	12.88	12.89	12.89	12.90	12.91	12.91	12.92
1.14	5.08	12.00	12.37	12.51	12.57	12.61	12.63	12.65	12.66	12.67	12.67	12.68	12.69	12.70	12.70	12.70
1.15	0.29	11.67	12.10	12.27	12.34	12.39	12.42	12.44	12.45	12.46	12.46	12.47	12.48	12.49	12.49	12.50
1.16	0.00	11.33	11.83	12.03	12.13	12.18	12.21	12.22	12.24	12.25	12.25	12.26	12.26	12.28	12.29	12.29
1.17	0.00	11.00	11.56	11.79	11.90	11.96	12.00	12.02	12.03	12.04	12.05	12.06	12.07	12.08	12.08	12.09
1.18	0.00	10.67	11.29	11.56	11.68	11.75	11.79	11.81	11.82	11.84	11.84	11.85	11.87	11.88	11.88	11.89
1.19	0.00	10.33	11.02	11.33	11.46	11.54	11.58	11.61	11.62	11.63	11.64	11.65	11.67	11.68	11.69	11.69
1.20	0.00	10.00	10.76	11.10	11.24	11.34	11.38	11.41	11.42	11.43	11.44	11.46	11.47	11.48	11.49	11.49
1.21	0.00	9.67	10.50	10.87	11.03	11.13	11.18	11.21	11.22	11.24	11.25	11.26	11.28	11.29	11.30	11.30
1.22	0.00	9.33	10.23	10.65	10.82	10.93	10.98	11.01	11.03	11.04	11.05	11.07	11.09	11.09	11.10	11.11
1.23	0.00	9.00	9.97	10.42	10.61	10.73	10.78	10.81	10.84	10.85	10.86	10.88	10.90	10.91	10.91	10.92
1.24	0.00	8.67	9.72	10.20	10.41	10.53	10.59	10.62	10.64	10.66	10.67	10.69	10.71	10.72	10.73	10.73
1.25	0.00	8.33	9.46	9.98	10.21	10.34	10.40	10.43	10.46	10.47	10.48	10.50	10.52	10.53	10.54	10.55
1.26	0.00	8.00	9.21	9.77	10.00	10.15	10.21	10.25	10.27	10.29	10.30	10.32	10.34	10.35	10.36	10.37
1.27	0.00	7.67	8.96	9.55	9.81	9.96	10.02	10.06	10.09	10.10	10.12	10.13	10.16	10.17	10.18	10.19
1.28	0.00	7.33	8.71	9.34	9.61	9.77	9.84	9.88	9.90	9.92	9.94	9.95	9.98	9.99	10.00	10.01
1.29	0.00	7.00	8.46	9.13	9.42	9.58	9.65	9.70	9.72	9.74	9.76	9.78	9.80	9.82	9.83	9.83
1.30	0.00	6.67	8.21	8.93	9.22	9.40	9.48	9.52	9.55	9.57	9.58	9.60	9.63	9.64	9.65	9.66
1.31	0.00	6.33	7.97	8.72	9.03	9.22	9.30	9.34	9.37	9.39	9.42	9.43	9.46	9.47	9.48	9.49
1.32	0.00	6.00	7.73	8.52	8.85	9.04	9.12	9.17	9.20	9.22	9.24	9.26	9.29	9.30	9.31	9.32
1.33	0.00	5.67	7.49	8.32	8.66	8.86	8.95	9.00	9.03	9.05	9.07	9.09	9.12	9.13	9.15	9.15
1.34	0.00	5.33	7.25	8.12	8.48	8.69	8.78	8.83	8.86	8.88	8.90	8.92	8.95	8.97	8.98	8.99
1.35	0.00	5.00	7.02	7.92	8.30	8.52	8.61	8.66	8.69	8.72	8.74	8.76	8.79	8.81	8.82	8.83
1.36	0.00	4.67	6.79	7.73	8.12	8.35	8.44	8.50	8.53	8.55	8.57	8.60	8.63	8.65	8.66	8.67
1.37	0.00	4.33	6.56	7.54	7.95	8.18	8.28	8.33	8.37	8.39	8.41	8.44	8.47	8.49	8.50	8.51
1.38	0.00	4.00	6.33	7.35	7.77	8.02	8.12	8.17	8.21	8.24	8.25	8.28	8.31	8.33	8.35	8.35
1.39	0.00	3.67	6.10	7.17	7.60	7.85	7.96	8.01	8.05	8.08	8.10	8.12	8.16	8.18	8.19	8.20
1.40	0.00	3.33	5.88	6.98	7.44	7.69	7.80	7.86	7.90	7.92	7.94	7.97	8.01	8.02	8.04	8.05
1.41	0.00	3.00	5.66	6.80	7.27	7.53	7.64	7.70	7.74	7.77	7.79	7.82	7.86	7.87	7.89	7.90
1.42	0.00	2.67	5.44	6.62	7.10	7.37	7.49	7.55	7.59	7.62	7.64	7.67	7.71	7.73	7.74	7.75
1.43	0.00	2.33	5.23	6.45	6.94	7.22	7.34	7.40	7.44	7.47	7.50	7.52	7.56	7.58	7.60	7.61
1.44	0.00	2.00	5.01	6.27	6.78	7.07	7.19	7.26	7.30	7.33	7.35	7.38	7.42	7.44	7.46	7.47
1.45	0.00	1.67	4.81	6.10	6.63	6.92	7.04	7.11	7.15	7.18	7.21	7.24	7.28	7.30	7.31	7.33
1.46	0.00	1.33	4.60	5.93	6.47	6.77	6.90	6.97	7.01	7.04	7.07	7.10	7.14	7.16	7.18	7.19
1.47	0.00	1.00	4.39	5.77	6.32	6.63	6.75	6.83	6.87	6.90	6.93	6.96	7.00	7.02	7.04	7.05
1.48	0.00	.67	4.19	5.60	6.17	6.48	6.61	6.69	6.73	6.77	6.79	6.82	6.86	6.88	6.90	6.91
1.49	0.00	.33	3.99	5.44	6.02	6.34	6.48	6.55	6.60	6.63	6.65	6.69	6.73	6.75	6.77	6.78

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1.50	0.00	0.00	3.80	5.28	5.87	6.20	6.34	6.41	6.46	6.50	6.52	6.55	6.60	6.62	6.64	6.65
1.51	0.00	0.00	3.61	5.13	5.73	6.06	6.20	6.28	6.33	6.36	6.39	6.42	6.47	6.49	6.51	6.52
1.52	0.00	0.00	3.42	4.97	5.59	5.93	6.07	6.15	6.20	6.23	6.26	6.29	6.34	6.36	6.38	6.39
1.53	0.00	0.00	3.23	4.82	5.45	5.80	5.94	6.02	6.07	6.11	6.13	6.17	6.21	6.24	6.26	6.27
1.54	0.00	0.00	3.05	4.67	5.31	5.67	5.81	5.89	5.95	5.98	6.01	6.04	6.09	6.11	6.13	6.15
1.55	0.00	0.00	2.87	4.52	5.18	5.54	5.69	5.77	5.82	5.86	5.88	5.92	5.97	5.99	6.01	6.02
1.56	0.00	0.00	2.69	4.38	5.05	5.41	5.56	5.65	5.70	5.74	5.76	5.80	5.85	5.87	5.89	5.90
1.57	0.00	0.00	2.52	4.24	4.92	5.29	5.44	5.53	5.58	5.62	5.64	5.68	5.73	5.75	5.78	5.79
1.58	0.00	0.00	2.35	4.10	4.79	5.16	5.32	5.41	5.46	5.50	5.53	5.56	5.61	5.64	5.66	5.67
1.59	0.00	0.00	2.19	3.96	4.66	5.04	5.20	5.29	5.34	5.38	5.42	5.45	5.50	5.52	5.54	5.56
1.60	0.00	0.00	2.03	3.83	4.54	4.92	5.09	5.17	5.23	5.27	5.30	5.33	5.38	5.41	5.43	5.44
1.61	0.00	0.00	1.87	3.69	4.41	4.81	4.97	5.06	5.12	5.16	5.18	5.22	5.27	5.30	5.32	5.33
1.62	0.00	0.00	1.72	3.57	4.30	4.69	4.86	4.95	5.01	5.04	5.07	5.11	5.16	5.19	5.21	5.23
1.63	0.00	0.00	1.57	3.44	4.18	4.58	4.75	4.84	4.90	4.94	4.97	5.01	5.06	5.08	5.11	5.12
1.64	0.00	0.00	1.42	3.31	4.06	4.47	4.64	4.73	4.79	4.83	4.86	4.90	4.95	4.98	5.00	5.01
1.65	0.00	0.00	1.28	3.19	3.95	4.36	4.53	4.62	4.68	4.72	4.75	4.79	4.85	4.87	4.90	4.91
1.66	0.00	0.00	1.15	3.07	3.84	4.25	4.43	4.52	4.58	4.62	4.65	4.69	4.74	4.77	4.80	4.81
1.67	0.00	0.00	1.02	2.95	3.73	4.15	4.32	4.42	4.48	4.52	4.55	4.59	4.64	4.67	4.70	4.71
1.68	0.00	0.00	0.89	2.84	3.62	4.05	4.22	4.32	4.38	4.42	4.45	4.49	4.55	4.57	4.60	4.61
1.69	0.00	0.00	0.77	2.73	3.52	3.94	4.12	4.22	4.28	4.32	4.35	4.39	4.45	4.47	4.50	4.51
1.70	0.00	0.00	0.66	2.62	3.41	3.84	4.02	4.12	4.18	4.22	4.25	4.30	4.35	4.38	4.41	4.42
1.71	0.00	0.00	0.55	2.51	3.31	3.75	3.93	4.02	4.09	4.13	4.16	4.20	4.26	4.29	4.31	4.32
1.72	0.00	0.00	0.45	2.41	3.21	3.65	3.83	3.93	3.99	4.04	4.07	4.11	4.17	4.19	4.22	4.23
1.73	0.00	0.00	0.36	2.30	3.11	3.56	3.74	3.84	3.90	3.94	3.98	4.02	4.08	4.10	4.13	4.14
1.74	0.00	0.00	0.27	2.20	3.02	3.46	3.65	3.75	3.81	3.85	3.89	3.93	3.99	4.01	4.04	4.05
1.75	0.00	0.00	0.19	2.11	2.93	3.37	3.56	3.66	3.72	3.77	3.80	3.84	3.90	3.93	3.95	3.97
1.76	0.00	0.00	0.12	2.01	2.83	3.28	3.47	3.57	3.63	3.68	3.71	3.76	3.81	3.84	3.87	3.88
1.77	0.00	0.00	0.06	1.92	2.74	3.20	3.38	3.48	3.55	3.59	3.63	3.67	3.73	3.76	3.78	3.80
1.78	0.00	0.00	0.02	1.83	2.66	3.11	3.30	3.40	3.47	3.51	3.54	3.59	3.64	3.67	3.70	3.71
1.79	0.00	0.00	0.00	1.74	2.57	3.03	3.21	3.32	3.38	3.43	3.46	3.51	3.56	3.59	3.62	3.63
1.80	0.00	0.00	0.00	1.65	2.49	2.94	3.13	3.24	3.30	3.35	3.38	3.43	3.48	3.51	3.54	3.55
1.81	0.00	0.00	0.00	1.57	2.40	2.86	3.05	3.16	3.22	3.27	3.30	3.35	3.40	3.43	3.46	3.47
1.82	0.00	0.00	0.00	1.49	2.32	2.79	2.98	3.08	3.15	3.19	3.22	3.27	3.33	3.36	3.38	3.40
1.83	0.00	0.00	0.00	1.41	2.25	2.71	2.90	3.00	3.07	3.11	3.15	3.19	3.25	3.28	3.31	3.32
1.84	0.00	0.00	0.00	1.34	2.17	2.63	2.82	2.93	2.99	3.04	3.07	3.12	3.18	3.21	3.23	3.25
1.85	0.00	0.00	0.00	1.26	2.09	2.56	2.75	2.85	2.92	2.97	3.00	3.05	3.10	3.13	3.16	3.17
1.86	0.00	0.00	0.00	1.19	2.02	2.48	2.68	2.78	2.85	2.89	2.93	2.97	3.03	3.06	3.09	3.10
1.87	0.00	0.00	0.00	1.12	1.95	2.41	2.61	2.71	2.78	2.82	2.86	2.90	2.96	2.99	3.02	3.03
1.88	0.00	0.00	0.00	1.06	1.88	2.34	2.54	2.64	2.71	2.75	2.79	2.83	2.89	2.92	2.95	2.96
1.89	0.00	0.00	0.00	0.99	1.81	2.28	2.47	2.57	2.64	2.69	2.72	2.77	2.83	2.85	2.88	2.90

TABLE IV

TABLE FOR ESTIMATING THE LOT PERCENTAGE FOR PLANS BASED ON UNKNOWN STANDARD DEVIATION—(*Continued*)

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1.90	0.00	0.00	0.00	0.93	1.75	2.21	2.40	2.51	2.57	2.62	2.65	2.70	2.76	2.79	2.82	2.81
1.91	0.00	0.00	0.00	0.87	1.68	2.14	2.34	2.44	2.51	2.56	2.59	2.63	2.69	2.72	2.75	2.77
1.92	0.00	0.00	0.00	0.81	1.62	2.08	2.27	2.38	2.45	2.49	2.52	2.57	2.63	2.66	2.69	2.70
1.93	0.00	0.00	0.00	0.76	1.56	2.02	2.21	2.32	2.38	2.43	2.46	2.51	2.57	2.60	2.62	2.64
1.94	0.00	0.00	0.00	0.70	1.50	1.96	2.15	2.25	2.32	2.37	2.40	2.45	2.51	2.54	2.56	2.58
1.95	0.00	0.00	0.00	0.65	1.44	1.90	2.09	2.19	2.26	2.31	2.34	2.39	2.45	2.48	2.50	2.52
1.96	0.00	0.00	0.00	0.60	1.38	1.84	2.03	2.14	2.20	2.25	2.28	2.33	2.39	2.42	2.44	2.46
1.97	0.00	0.00	0.00	0.56	1.33	1.78	1.97	2.08	2.14	2.19	2.22	2.27	2.33	2.36	2.39	2.40
1.98	0.00	0.00	0.00	0.51	1.27	1.73	1.92	2.02	2.09	2.13	2.17	2.21	2.27	2.30	2.33	2.34
1.99	0.00	0.00	0.00	0.47	1.22	1.67	1.86	1.97	2.03	2.08	2.11	2.16	2.22	2.25	2.27	2.29
2.00	0.00	0.00	0.00	0.43	1.17	1.62	1.81	1.91	1.98	2.03	2.06	2.10	2.16	2.19	2.22	2.23
2.01	0.00	0.00	0.00	0.39	1.12	1.57	1.76	1.86	1.93	1.97	2.01	2.05	2.11	2.14	2.17	2.18
2.02	0.00	0.00	0.00	0.36	1.07	1.52	1.71	1.81	1.87	1.92	1.95	2.00	2.06	2.09	2.11	2.13
2.03	0.00	0.00	0.00	0.32	1.03	1.47	1.66	1.76	1.82	1.87	1.90	1.95	2.01	2.04	2.06	2.08
2.04	0.00	0.00	0.00	0.29	0.98	1.42	1.61	1.71	1.77	1.82	1.85	1.90	1.96	1.99	2.01	2.03
2.05	0.00	0.00	0.00	0.26	0.94	1.37	1.56	1.66	1.73	1.77	1.80	1.85	1.91	1.94	1.96	1.98
2.06	0.00	0.00	0.00	0.23	0.90	1.33	1.51	1.61	1.68	1.72	1.76	1.80	1.86	1.89	1.92	1.93
2.07	0.00	0.00	0.00	0.21	0.86	1.28	1.47	1.57	1.63	1.68	1.71	1.76	1.81	1.84	1.87	1.88
2.08	0.00	0.00	0.00	0.18	0.82	1.24	1.42	1.52	1.59	1.63	1.66	1.71	1.77	1.79	1.82	1.84
2.09	0.00	0.00	0.00	0.16	0.78	1.20	1.38	1.48	1.54	1.59	1.62	1.66	1.72	1.75	1.78	1.79
2.10	0.00	0.00	0.00	0.14	0.74	1.16	1.34	1.44	1.50	1.54	1.58	1.62	1.68	1.71	1.73	1.75
2.11	0.00	0.00	0.00	0.12	0.71	1.12	1.30	1.39	1.46	1.50	1.53	1.58	1.63	1.66	1.69	1.70
2.12	0.00	0.00	0.00	0.10	0.67	1.08	1.26	1.35	1.42	1.46	1.49	1.54	1.59	1.62	1.65	1.66
2.13	0.00	0.00	0.00	0.08	0.64	1.04	1.22	1.31	1.38	1.42	1.45	1.50	1.55	1.58	1.61	1.62
2.14	0.00	0.00	0.00	0.07	0.61	1.00	1.18	1.28	1.34	1.38	1.41	1.46	1.51	1.54	1.57	1.58
2.15	0.00	0.00	0.00	0.06	0.58	0.97	1.14	1.24	1.30	1.34	1.37	1.42	1.47	1.50	1.53	1.54
2.16	0.00	0.00	0.00	0.05	0.55	0.93	1.10	1.20	1.26	1.30	1.34	1.38	1.43	1.46	1.49	1.50
2.17	0.00	0.00	0.00	0.04	0.52	0.90	1.07	1.16	1.22	1.27	1.30	1.34	1.40	1.42	1.45	1.46
2.18	0.00	0.00	0.00	0.03	0.49	0.87	1.03	1.13	1.19	1.23	1.26	1.30	1.36	1.39	1.41	1.42
2.19	0.00	0.00	0.00	0.02	0.46	0.83	1.00	1.09	1.15	1.20	1.23	1.27	1.32	1.35	1.38	1.39
2.20	0.000	0.000	0.000	0.015	0.437	0.803	0.968	1.061	1.120	1.161	1.192	1.233	1.287	1.314	1.340	1.352
2.21	0.000	0.000	0.000	0.010	0.413	0.772	0.936	1.028	1.087	1.128	1.158	1.199	1.253	1.279	1.305	1.318
2.22	0.000	0.000	0.000	0.006	0.389	0.743	0.905	0.996	1.054	1.095	1.125	1.166	1.219	1.245	1.271	1.283
2.23	0.000	0.000	0.000	0.003	0.366	0.715	0.875	0.965	1.023	1.063	1.093	1.134	1.186	1.212	1.238	1.250
2.24	0.000	0.000	0.000	0.002	0.345	0.687	0.845	0.935	0.992	1.032	1.061	1.102	1.154	1.180	1.205	1.218
2.25	0.000	0.000	0.000	0.001	0.324	0.660	0.816	0.905	0.962	1.002	1.031	1.071	1.123	1.148	1.173	1.186
2.26	0.000	0.000	0.000	0.000	0.304	0.634	0.789	0.876	0.933	0.972	1.001	1.041	1.092	1.117	1.142	1.155
2.27	0.000	0.000	0.000	0.000	0.285	0.609	0.762	0.848	0.904	0.943	0.972	1.011	1.062	1.087	1.112	1.124
2.28	0.000	0.000	0.000	0.000	0.267	0.585	0.735	0.821	0.876	0.915	0.943	0.982	1.033	1.058	1.082	1.094
2.29	0.000	0.000	0.000	0.000	0.250	0.561	0.710	0.794	0.849	0.887	0.915	0.954	1.004	1.029	1.053	1.065

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
2.30	0.000	0.000	0.000	0.000	0.233	0.538	0.685	0.769	0.823	0.861	0.888	0.927	0.977	1.001	1.025	1.037
2.31	0.000	0.000	0.000	0.000	0.218	0.516	0.661	0.743	0.797	0.834	0.862	0.900	0.949	0.974	0.997	1.009
2.32	0.000	0.000	0.000	0.000	0.203	0.495	0.637	0.719	0.772	0.809	0.836	0.874	0.923	0.947	0.971	0.982
2.33	0.000	0.000	0.000	0.000	0.189	0.474	0.614	0.695	0.748	0.784	0.811	0.848	0.897	0.921	0.944	0.956
2.34	0.000	0.000	0.000	0.000	0.175	0.454	0.592	0.672	0.724	0.760	0.787	0.824	0.872	0.895	0.915	0.930
2.35	0.000	0.000	0.000	0.000	0.163	0.435	0.571	0.650	0.701	0.736	0.763	0.799	0.847	0.870	0.893	0.905
2.36	0.000	0.000	0.000	0.000	0.151	0.416	0.550	0.628	0.678	0.714	0.740	0.776	0.823	0.846	0.869	0.880
2.37	0.000	0.000	0.000	0.000	0.139	0.398	0.530	0.606	0.656	0.691	0.717	0.753	0.799	0.822	0.845	0.856
2.38	0.000	0.000	0.000	0.000	0.128	0.381	0.510	0.586	0.635	0.670	0.695	0.730	0.777	0.799	0.822	0.833
2.39	0.000	0.000	0.000	0.000	0.118	0.364	0.491	0.566	0.614	0.648	0.674	0.709	0.754	0.777	0.799	0.810
2.40	0.000	0.000	0.000	0.000	0.109	0.348	0.473	0.546	0.594	0.628	0.653	0.687	0.732	0.755	0.777	0.787
2.41	0.000	0.000	0.000	0.000	0.100	0.332	0.455	0.527	0.575	0.608	0.633	0.667	0.711	0.733	0.755	0.766
2.42	0.000	0.000	0.000	0.000	0.091	0.317	0.437	0.509	0.555	0.588	0.613	0.646	0.691	0.712	0.734	0.744
2.43	0.000	0.000	0.000	0.000	0.083	0.302	0.421	0.491	0.537	0.569	0.593	0.627	0.670	0.692	0.713	0.724
2.44	0.000	0.000	0.000	0.000	0.076	0.288	0.404	0.474	0.519	0.551	0.575	0.608	0.651	0.672	0.693	0.703
2.45	0.000	0.000	0.000	0.000	0.069	0.275	0.389	0.457	0.501	0.533	0.556	0.589	0.632	0.653	0.673	0.684
2.46	0.000	0.000	0.000	0.000	0.063	0.262	0.373	0.440	0.484	0.516	0.539	0.571	0.613	0.634	0.654	0.664
2.47	0.000	0.000	0.000	0.000	0.057	0.249	0.359	0.425	0.468	0.499	0.521	0.553	0.595	0.615	0.636	0.646
2.48	0.000	0.000	0.000	0.000	0.051	0.237	0.344	0.409	0.452	0.482	0.505	0.536	0.577	0.597	0.617	0.627
2.49	0.000	0.000	0.000	0.000	0.046	0.226	0.331	0.394	0.436	0.466	0.488	0.519	0.560	0.580	0.600	0.609
2.50	0.000	0.000	0.000	0.000	0.041	0.214	0.317	0.380	0.421	0.451	0.473	0.503	0.543	0.563	0.582	0.592
2.51	0.000	0.000	0.000	0.000	0.037	0.204	0.304	0.366	0.407	0.436	0.457	0.487	0.527	0.546	0.565	0.575
2.52	0.000	0.000	0.000	0.000	0.033	0.193	0.292	0.352	0.392	0.421	0.442	0.472	0.511	0.530	0.549	0.558
2.53	0.000	0.000	0.000	0.000	0.029	0.184	0.280	0.339	0.379	0.407	0.428	0.457	0.495	0.514	0.533	0.542
2.54	0.000	0.000	0.000	0.000	0.026	0.174	0.268	0.326	0.365	0.393	0.413	0.442	0.480	0.499	0.517	0.527
2.55	0.000	0.000	0.000	0.000	0.023	0.165	0.257	0.314	0.352	0.379	0.400	0.428	0.465	0.484	0.502	0.511
2.56	0.000	0.000	0.000	0.000	0.020	0.156	0.246	0.302	0.340	0.366	0.386	0.414	0.451	0.469	0.487	0.496
2.57	0.000	0.000	0.000	0.000	0.017	0.148	0.236	0.291	0.327	0.354	0.373	0.401	0.437	0.455	0.473	0.482
2.58	0.000	0.000	0.000	0.000	0.015	0.140	0.226	0.279	0.316	0.341	0.361	0.388	0.424	0.441	0.459	0.468
2.59	0.000	0.000	0.000	0.000	0.013	0.133	0.216	0.269	0.304	0.330	0.349	0.375	0.410	0.428	0.445	0.454
2.60	0.000	0.000	0.000	0.000	0.011	0.125	0.207	0.258	0.293	0.318	0.337	0.363	0.398	0.415	0.432	0.441
2.61	0.000	0.000	0.000	0.000	0.009	0.118	0.198	0.248	0.282	0.307	0.325	0.351	0.385	0.402	0.419	0.428
2.62	0.000	0.000	0.000	0.000	0.008	0.112	0.189	0.238	0.272	0.296	0.314	0.339	0.373	0.390	0.406	0.415
2.63	0.000	0.000	0.000	0.000	0.007	0.105	0.181	0.229	0.262	0.285	0.303	0.328	0.361	0.378	0.394	0.402
2.64	0.000	0.000	0.000	0.000	0.006	0.099	0.172	0.220	0.252	0.275	0.293	0.317	0.350	0.366	0.382	0.390
2.65	0.000	0.000	0.000	0.000	0.005	0.094	0.165	0.211	0.243	0.265	0.282	0.307	0.339	0.355	0.371	0.379
2.66	0.000	0.000	0.000	0.000	0.004	0.088	0.157	0.202	0.233	0.256	0.273	0.296	0.328	0.344	0.359	0.367
2.67	0.000	0.000	0.000	0.000	0.003	0.083	0.150	0.194	0.224	0.246	0.263	0.286	0.317	0.333	0.348	0.356
2.68	0.000	0.000	0.000	0.000	0.002	0.078	0.143	0.186	0.216	0.237	0.254	0.277	0.307	0.322	0.338	0.345
2.69	0.000	0.000	0.000	0.000	0.002	0.073	0.136	0.179	0.208	0.229	0.245	0.267	0.297	0.312	0.327	0.335

TABLE IV

TABLE FOR ESTIMATING THE LOT PERCENTAGE FOR PLANS BASED ON UNKNOWN STANDARD DEVIATION—(*Continued*)

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
2.70	0.000	0.000	0.000	0.000	0.001	0.069	0.130	0.171	0.200	0.220	0.236	0.258	0.288	0.302	0.317	0.323
2.71	0.000	0.000	0.000	0.000	0.001	0.064	0.124	0.164	0.192	0.212	0.227	0.249	0.276	0.293	0.307	0.315
2.72	0.000	0.000	0.000	0.000	0.000	0.060	0.118	0.157	0.184	0.204	0.219	0.241	0.269	0.283	0.298	0.305
2.73	0.000	0.000	0.000	0.000	0.000	0.057	0.112	0.151	0.177	0.197	0.211	0.232	0.260	0.274	0.288	0.296
2.74	0.000	0.000	0.000	0.000	0.000	0.053	0.107	0.144	0.170	0.189	0.204	0.224	0.252	0.266	0.279	0.286
2.75	0.000	0.000	0.000	0.000	0.000	0.049	0.102	0.138	0.163	0.182	0.196	0.216	0.243	0.257	0.271	0.277
2.76	0.000	0.000	0.000	0.000	0.000	0.046	0.097	0.132	0.157	0.175	0.189	0.209	0.235	0.249	0.262	0.269
2.77	0.000	0.000	0.000	0.000	0.000	0.043	0.092	0.126	0.151	0.168	0.182	0.201	0.227	0.241	0.254	0.260
2.78	0.000	0.000	0.000	0.000	0.000	0.040	0.087	0.121	0.145	0.162	0.175	0.194	0.220	0.233	0.246	0.252
2.79	0.000	0.000	0.000	0.000	0.000	0.037	0.083	0.115	0.139	0.156	0.169	0.187	0.212	0.225	0.238	0.244
2.80	0.000	0.000	0.000	0.000	0.000	0.035	0.079	0.110	0.133	0.150	0.162	0.181	0.205	0.218	0.230	0.237
2.81	0.000	0.000	0.000	0.000	0.000	0.032	0.075	0.105	0.128	0.144	0.156	0.174	0.198	0.211	0.223	0.229
2.82	0.000	0.000	0.000	0.000	0.000	0.030	0.071	0.101	0.122	0.138	0.150	0.168	0.192	0.204	0.216	0.222
2.83	0.000	0.000	0.000	0.000	0.000	0.028	0.067	0.096	0.117	0.133	0.145	0.162	0.185	0.197	0.209	0.215
2.84	0.000	0.000	0.000	0.000	0.000	0.026	0.064	0.092	0.112	0.128	0.139	0.156	0.179	0.190	0.202	0.208
2.85	0.000	0.000	0.000	0.000	0.000	0.024	0.060	0.088	0.108	0.122	0.134	0.150	0.173	0.184	0.195	0.201
2.86	0.000	0.000	0.000	0.000	0.000	0.022	0.057	0.084	0.103	0.118	0.129	0.145	0.167	0.178	0.189	0.195
2.87	0.000	0.000	0.000	0.000	0.000	0.020	0.054	0.080	0.099	0.113	0.124	0.139	0.161	0.172	0.183	0.188
2.88	0.000	0.000	0.000	0.000	0.000	0.019	0.051	0.076	0.094	0.108	0.119	0.134	0.155	0.166	0.177	0.182
2.89	0.000	0.000	0.000	0.000	0.000	0.017	0.048	0.073	0.090	0.104	0.114	0.129	0.150	0.160	0.171	0.176
2.90	0.000	0.000	0.000	0.000	0.000	0.016	0.046	0.069	0.087	0.100	0.110	0.125	0.145	0.155	0.165	0.171
2.91	0.000	0.000	0.000	0.000	0.000	0.015	0.043	0.066	0.083	0.096	0.106	0.120	0.140	0.150	0.160	0.165
2.92	0.000	0.000	0.000	0.000	0.000	0.013	0.041	0.063	0.079	0.092	0.101	0.115	0.135	0.145	0.155	0.160
2.93	0.000	0.000	0.000	0.000	0.000	0.012	0.038	0.060	0.076	0.088	0.097	0.111	0.130	0.140	0.149	0.154
2.94	0.000	0.000	0.000	0.000	0.000	0.011	0.036	0.057	0.072	0.084	0.093	0.107	0.125	0.135	0.144	0.149
2.95	0.000	0.000	0.000	0.000	0.000	0.010	0.034	0.054	0.069	0.081	0.090	0.103	0.121	0.130	0.140	0.144
2.96	0.000	0.000	0.000	0.000	0.000	0.009	0.032	0.051	0.066	0.077	0.086	0.099	0.117	0.126	0.135	0.140
2.97	0.000	0.000	0.000	0.000	0.000	0.009	0.030	0.049	0.063	0.074	0.083	0.095	0.112	0.121	0.130	0.135
2.98	0.000	0.000	0.000	0.000	0.000	0.008	0.028	0.046	0.060	0.071	0.079	0.092	0.108	0.117	0.126	0.130
2.99	0.000	0.000	0.000	0.000	0.000	0.007	0.027	0.044	0.057	0.068	0.076	0.088	0.104	0.113	0.122	0.126
3.00	0.000	0.000	0.000	0.000	0.000	0.006	0.025	0.042	0.055	0.065	0.073	0.084	0.101	0.109	0.118	0.122
3.01	0.000	0.000	0.000	0.000	0.000	0.006	0.024	0.040	0.052	0.062	0.070	0.081	0.097	0.105	0.114	0.118
3.02	0.000	0.000	0.000	0.000	0.000	0.005	0.022	0.038	0.050	0.059	0.067	0.078	0.093	0.101	0.110	0.114
3.03	0.000	0.000	0.000	0.000	0.000	0.005	0.021	0.036	0.048	0.057	0.064	0.075	0.090	0.098	0.106	0.110
3.04	0.000	0.000	0.000	0.000	0.000	0.004	0.019	0.034	0.045	0.054	0.061	0.072	0.087	0.094	0.102	0.106
3.05	0.000	0.000	0.000	0.000	0.000	0.004	0.018	0.032	0.043	0.052	0.059	0.069	0.083	0.091	0.099	0.103
3.06	0.000	0.000	0.000	0.000	0.000	0.003	0.017	0.030	0.041	0.050	0.056	0.066	0.080	0.088	0.095	0.099
3.07	0.000	0.000	0.000	0.000	0.000	0.003	0.016	0.029	0.039	0.047	0.054	0.064	0.077	0.085	0.092	0.096
3.08	0.000	0.000	0.000	0.000	0.000	0.003	0.015	0.027	0.037	0.045	0.052	0.061	0.074	0.082	0.089	0.092
3.09	0.000	0.000	0.000	0.000	0.000	0.002	0.014	0.026	0.036	0.043	0.049	0.059	0.072	0.079	0.086	0.089

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
3.10	0.000	0.000	0.000	0.000	0.000	0.001	0.013	0.024	0.034	0.042	0.047	0.056	0.069	0.076	0.083	0.086
3.11	0.000	0.000	0.000	0.000	0.000	0.002	0.012	0.023	0.032	0.039	0.045	0.054	0.066	0.073	0.080	0.083
3.12	0.000	0.000	0.000	0.000	0.000	0.002	0.011	0.022	0.031	0.038	0.043	0.052	0.064	0.070	0.077	0.080
3.13	0.000	0.000	0.000	0.000	0.000	0.002	0.011	0.021	0.029	0.036	0.042	0.050	0.061	0.068	0.074	0.077
3.14	0.000	0.000	0.000	0.000	0.000	0.001	0.010	0.019	0.028	0.034	0.040	0.048	0.059	0.065	0.071	0.075
3.15	0.000	0.000	0.000	0.000	0.000	0.001	0.009	0.018	0.026	0.033	0.038	0.046	0.057	0.063	0.069	0.072
3.16	0.000	0.000	0.000	0.000	0.000	0.001	0.009	0.017	0.025	0.032	0.036	0.044	0.055	0.060	0.066	0.069
3.17	0.000	0.000	0.000	0.000	0.000	0.001	0.008	0.016	0.024	0.030	0.035	0.042	0.053	0.058	0.064	0.067
3.18	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.015	0.022	0.028	0.033	0.040	0.050	0.056	0.062	0.065
3.19	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.015	0.021	0.027	0.032	0.038	0.049	0.054	0.059	0.062
3.20	0.000	0.000	0.000	0.000	0.000	0.001	0.006	0.014	0.020	0.026	0.030	0.037	0.047	0.052	0.057	0.060
3.21	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.013	0.019	0.024	0.029	0.035	0.045	0.050	0.055	0.058
3.22	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.012	0.018	0.023	0.027	0.034	0.043	0.048	0.053	0.056
3.23	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.011	0.017	0.022	0.026	0.032	0.041	0.046	0.051	0.054
3.24	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.011	0.016	0.021	0.025	0.031	0.040	0.044	0.049	0.052
3.25	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.010	0.015	0.020	0.024	0.030	0.038	0.043	0.048	0.050
3.26	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.009	0.015	0.019	0.023	0.028	0.037	0.041	0.046	0.048
3.27	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.009	0.014	0.019	0.022	0.027	0.035	0.040	0.044	0.046
3.28	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.008	0.013	0.017	0.021	0.026	0.034	0.038	0.042	0.045
3.29	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.008	0.012	0.016	0.020	0.025	0.032	0.037	0.041	0.043
3.30	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.007	0.012	0.015	0.019	0.024	0.031	0.035	0.039	0.042
3.31	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.007	0.011	0.015	0.018	0.023	0.030	0.034	0.038	0.040
3.32	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.006	0.010	0.014	0.017	0.022	0.029	0.032	0.036	0.039
3.33	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.006	0.010	0.013	0.016	0.021	0.027	0.031	0.035	0.037
3.34	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.006	0.009	0.013	0.015	0.020	0.026	0.030	0.034	0.036
3.35	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.005	0.009	0.012	0.015	0.019	0.025	0.029	0.032	0.034
3.36	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.005	0.008	0.011	0.014	0.018	0.024	0.028	0.031	0.033
3.37	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.005	0.008	0.011	0.013	0.017	0.023	0.026	0.030	0.032
3.38	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.007	0.010	0.013	0.016	0.022	0.025	0.029	0.031
3.39	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.007	0.010	0.012	0.016	0.021	0.024	0.028	0.030
3.40	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.007	0.009	0.011	0.015	0.020	0.023	0.027	0.029
3.41	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.006	0.009	0.011	0.014	0.020	0.022	0.026	0.027
3.42	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.006	0.008	0.010	0.014	0.019	0.021	0.025	0.026
3.43	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.005	0.008	0.010	0.013	0.018	0.021	0.024	0.025
3.44	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.005	0.007	0.009	0.012	0.017	0.020	0.023	0.024
3.45	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.005	0.007	0.009	0.012	0.016	0.019	0.022	0.023
3.46	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.005	0.007	0.008	0.012	0.016	0.018	0.021	0.022
3.47	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.006	0.008	0.011	0.015	0.017	0.020	0.021
3.48	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.006	0.007	0.010	0.014	0.017	0.019	0.021
3.49	0.000	0.000	0.000	0.000	0.000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

TABLE IV

TABLE FOR ESTIMATING THE LOT PERCENTAGE FOR PLANS BASED ON UNKNOWN STANDARD DEVIATION—(*Continued*)

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
3.50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.003	0.005	0.007	0.009	0.013	0.015	0.018	0.019
3.51	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.003	0.005	0.006	0.009	0.013	0.015	0.017	0.018
3.52	0.000	0.000	000	0.000	0.000	0.000	0.000	0.002	0.003	0.005	0.006	0.008	0.012	0.014	0.017	0.018
3.53	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.004	0.006	0.008	0.012	0.014	0.016	0.017
3.54	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.004	0.005	0.008	0.011	0.013	0.015	0.016
3.55	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.004	0.005	0.007	0.011	0.012	0.015	0.016
3.56	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.005	0.007	0.010	0.012	0.014	0.015
3.57	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.005	0.006	0.010	0.011	0.013	0.014
3.58	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.004	0.006	0.009	0.011	0.013	0.014
3.59	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.004	0.006	0.009	0.010	0.012	0.013
3.60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.004	0.006	0.008	0.010	0.012	0.013
3.61	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.004	0.005	0.008	0.010	0.011	0.012
3.62	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.003	0.005	0.008	0.009	0.011	0.012
3.63	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.007	0.009	0.010	0.011
3.64	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.007	0.008	0.010	0.011
3.65	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.007	0.008	0.010	0.010
3.66	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.004	0.006	0.008	0.009	0.010
3.67	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.004	0.006	0.007	0.009	0.010
3.68	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.002	0.004	0.006	0.007	0.008	0.009
3.69	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.002	0.003	0.005	0.007	0.008	0.009

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
3.70	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.002	0.003	0.005	0.006	0.008	0.008
3.71	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.006	0.007	0.008
3.72	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.006	0.007	0.008
3.73	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.006	0.007	0.007
3.74	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.005	0.007	0.007
3.75	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.002	0.004	0.005	0.006	0.007
3.76	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.002	0.004	0.005	0.006	0.007
3.77	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.002	0.004	0.005	0.006	0.006
3.78	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.004	0.004	0.005	0.006
3.79	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.005	0.006
3.80	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.005	0.006
3.81	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.005	0.005
3.82	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.005	0.005
3.83	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.004	0.005
3.84	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.003	0.003	0.004	0.005
3.85	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.002	0.003	0.004	0.004
3.86	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.004
3.87	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.004
3.88	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.004
3.89	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.003	0.004
3.90	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.003	0.004

TABLE V. MASTER TABLE FOR SAMPLING PLANS BASED ON THE AVERAGE RANGE

Sample size code letter	Sample size	h factor	Acceptable Quality Levels													
			.04	065	10	.15	25	.40	.65	1.00	1.50	2.50	4.00	6.50	10.00	15.00
			p^*	p^*	p^*	p^*	p^*	p^*	p^*	p^*	p^*	p^*	p^*	p^*	p^*	p^*
B	3	1.910										7.59	18.86	26.94	33.69	40.47
C	4	2.243								1.53	5.50	10.92	16.45	22.86	29.45	36.90
D	5	2.607							0.650	1.92	4.21	8.10	12.73	20.57	25.30	33.03
E	7	2.907					0.420	0.580	1.54	2.89	4.67	7.74	11.59	16.81	22.85	30.16
F	10	2.439				0.175	0.480	0.990	1.85	2.98	4.47	7.08	10.41	15.12	20.71	27.61
G	15	2.394	0.053	0.123	0.233	0.401	0.740	1.24	2.02	3.02	4.32	6.62	9.61	13.92	19.16	25.78
H	25	2.363	0.121	0.208	0.328	0.496	0.812	1.25	1.93	2.79	3.93	5.94	8.60	12.53	17.42	23.74
I	30	2.356	0.144	0.236	0.360	0.530	0.846	1.28	1.95	2.79	3.89	5.85	8.47	12.32	17.15	23.39
J	35	2.351	0.162	0.258	0.387	0.558	0.876	1.32	1.97	2.81	3.89	5.82	8.40	12.21	17.00	23.18
K	40	2.348	0.159	0.250	0.372	0.535	0.836	1.25	1.87	2.68	3.72	5.59	8.09	11.82	16.53	22.66
L	50	2.343	0.168	0.260	0.380	0.539	0.834	1.24	1.60	2.62	3.63	5.46	7.89	11.55	16.19	22.24
M	60	2.340	0.157	0.243	0.355	0.503	0.778	1.16	1.74	2.47	3.43	5.17	7.53	11.09	15.63	21.61
N	85	2.336	0.156	0.241	0.349	0.493	0.754	1.12	1.67	2.37	3.29	4.97	7.27	10.72	15.17	21.05
O	115	2.333	0.153	0.230	0.333	0.467	0.717	1.06	1.58	2.25	3.14	4.76	6.99	10.37	14.74	20.57
P	175	2.332	0.139	0.210	0.303	0.427	0.655	0.972	1.46	2.08	2.93	4.47	6.60	9.89	14.15	19.88
Q	230	2.330	0.142	0.215	0.308	0.432	0.661	0.976	1.47	2.08	2.92	4.46	6.57	9.84	14.10	19.82

All AQL and table values are in per cent defective

TABLE VI*

TABLE FOR ESTIMATING THE LOT PERCENTAGE FOR PLANS BASED ON AVERAGE RANGE

	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
0	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00	50.00
.1	47.24	46.67	46.44	46.29	46.20	46.13	46.08	46.07	46.06	46.05	46.05	46.04	46.03	46.03	46.02	46.02
.2	44.46	43.33	42.90	42.60	42.42	42.29	42.19	42.17	42.16	42.15	42.13	42.12	42.10	42.10	42.08	42.08
.3	41.63	40.00	39.37	38.95	38.70	38.51	38.38	38.34	38.32	38.31	38.28	38.27	38.25	38.24	38.23	38.22
.31	41.35	39.67	39.02	38.59	38.33	38.14	38.00	37.96	37.94	37.93	37.90	37.89	37.88	37.86	37.85	37.84
.32	41.08	39.33	38.67	38.23	37.96	37.77	37.63	37.59	37.57	37.55	37.53	37.51	37.50	37.48	37.47	37.46
.33	40.77	39.00	38.32	37.87	37.60	37.39	37.25	37.21	37.19	37.18	37.15	37.14	37.12	37.11	37.09	37.09
.34	40.49	38.67	37.97	37.51	37.23	37.02	36.88	36.84	36.82	36.80	36.77	36.76	36.74	36.73	36.72	36.71
.35	40.20	38.33	37.62	37.15	36.87	36.65	36.50	36.46	36.44	36.43	36.40	36.38	36.37	36.36	36.34	36.33
.36	39.91	38.00	37.28	36.79	36.50	36.29	36.13	36.09	36.07	36.05	36.03	36.01	35.99	35.97	35.96	35.96
.37	39.62	37.67	36.93	36.43	36.14	35.92	35.76	35.72	35.70	35.68	35.65	35.64	35.62	35.61	35.59	35.59
.38	39.33	37.33	36.58	36.07	35.78	35.55	35.39	35.35	35.33	35.31	35.28	35.27	35.25	35.24	35.23	35.22
.39	39.03	37.00	36.23	35.72	35.42	35.19	35.02	34.98	34.96	34.94	34.92	34.90	34.88	34.87	34.85	34.85
.40	38.74	36.67	35.88	35.36	35.05	34.82	34.66	34.62	34.59	34.58	34.55	34.53	34.51	34.49	34.48	34.48
.41	38.45	36.33	35.54	35.01	34.69	34.46	34.29	34.25	34.23	34.21	34.18	34.17	34.14	34.12	34.11	34.11
.42	38.15	36.00	35.19	34.65	34.33	34.10	33.93	33.89	33.86	33.85	33.82	33.80	33.78	33.77	33.75	33.74
.43	37.85	35.67	34.85	34.30	33.98	33.74	33.57	33.53	33.50	33.48	33.45	33.44	33.41	33.39	33.38	33.38
.44	37.56	35.33	34.50	33.95	33.62	33.38	33.21	33.17	33.14	33.12	33.09	33.08	33.05	33.03	33.02	33.02
.45	37.26	35.00	34.16	33.60	33.27	33.02	32.85	32.81	32.78	32.76	32.73	32.72	32.69	32.67	32.66	32.66
.46	36.96	34.67	33.81	33.24	32.91	32.66	32.49	32.45	32.42	32.40	32.37	32.36	32.33	32.31	32.30	32.30
.47	36.66	34.33	33.47	32.89	32.56	32.31	32.13	32.09	32.06	32.04	32.01	32.00	[illegible]	31.95	31.94	31.94
.48	36.35	34.00	33.12	32.55	32.21	31.96	31.78	31.74	31.71	31.69	31.66	31.64	31.62	31.61	31.59	31.58
.49	36.05	33.67	32.78	32.20	31.86	31.60	31.42	31.38	31.35	31.33	31.30	31.29	31.26	31.24	31.23	31.23
.50	35.75	33.33	32.44	31.85	31.51	31.25	31.07	31.03	31.00	30.98	30.95	30.94	30.91	30.89	30.88	30.87
.51	35.44	33.00	32.10	31.51	31.16	30.90	30.72	30.68	30.65	30.63	30.60	30.59	30.56	30.55	30.53	30.52
.52	35.13	32.67	31.76	31.16	30.81	30.55	30.37	30.33	30.30	30.28	30.25	30.24	30.21	30.19	30.18	30.17
.53	34.82	32.33	31.42	30.82	30.46	30.21	30.02	29.98	29.95	29.93	29.90	29.89	29.86	29.84	29.83	29.83
.54	34.51	32.00	31.08	30.47	30.12	29.86	29.68	29.64	29.61	29.59	29.56	29.54	29.52	29.50	29.48	29.48
.55	34.20	31.67	30.74	30.13	29.78	29.52	29.33	29.29	29.26	29.24	29.21	29.20	29.17	29.15	29.14	29.14
.56	33.88	31.33	30.40	29.79	29.44	29.18	28.99	28.95	28.92	28.90	28.87	28.86	28.83	28.81	28.80	28.79
.57	33.57	31.00	30.06	29.45	29.09	28.83	28.65	28.61	28.58	28.56	28.53	28.52	28.49	28.47	28.46	28.45
.58	33.25	30.67	29.73	29.11	28.76	28.50	28.31	28.27	28.24	28.22	28.19	28.18	28.15	28.13	28.12	28.12
.59	32.93	30.33	29.39	28.77	28.42	28.16	27.97	27.93	27.91	27.89	27.86	27.84	27.82	27.80	27.78	27.78
.60	32.61	30.00	29.05	28.44	28.08	27.82	27.64	27.60	27.57	27.55	27.52	27.51	27.48	27.46	27.45	27.45
.61	32.28	29.67	28.72	28.10	27.75	27.49	27.31	27.27	27.24	27.22	27.19	27.17	27.15	27.14	27.12	27.11
.62	31.96	29.33	28.39	27.77	27.41	27.16	26.97	26.93	26.91	26.89	26.86	26.84	26.82	26.81	26.79	26.78
.63	31.63	29.00	28.05	27.44	27.08	26.82	26.64	26.60	26.58	26.56	26.53	26.51	26.49	26.48	26.46	26.45
.64	31.30	28.67	27.72	27.11	26.75	26.50	26.32	26.28	26.25	26.23	26.20	26.19	26.16	26.14	26.13	26.13
.65	30.97	28.33	27.39	26.78	26.42	26.17	25.99	25.95	25.92	25.90	25.87	25.86	25.84	25.82	25.81	25.80
.66	30.63	28.00	27.06	26.45	26.10	25.84	25.67	25.63	25.60	25.58	25.55	25.54	25.52	25.50	25.48	25.48
.67	30.30	27.67	26.73	26.13	25.77	25.52	25.34	25.30	25.28	25.26	25.23	25.22	25.20	25.18	25.16	25.16
.68	29.96	27.33	26.40	25.79	25.45	25.20	25.02	24.98	24.96	24.94	24.91	24.90	24.88	24.87	24.85	24.84
.69	29.61	27.00	26.07	25.47	25.12	24.88	24.71	24.67	24.64	24.62	24.59	24.58	24.56	24.55	24.53	24.53
.70	29.27	26.67	25.74	25.14	24.80	24.56	24.39	24.35	24.33	24.31	24.28	24.27	24.25	24.24	24.22	24.22
.71	28.92	26.33	25.41	24.82	24.48	24.24	24.07	24.03	24.01	23.99	23.97	23.95	23.93	23.92	23.90	23.90
.72	28.57	26.00	25.09	24.50	24.17	23.93	23.76	23.72	23.70	23.68	23.66	23.64	23.62	23.60	23.59	23.59
.73	28.22	25.67	24.76	24.18	23.85	23.61	23.45	23.41	23.39	23.37	23.35	23.33	23.31	23.30	23.29	23.28
.74	27.86	25.33	24.44	23.86	23.54	23.30	23.14	23.10	23.08	23.07	23.04	23.03	23.01	22.99	22.98	22.98
.75	27.50	25.00	24.11	23.55	23.22	22.99	22.84	22.80	22.78	22.76	22.74	22.72	22.71	22.69	22.68	22.68
.76	27.13	24.67	23.79	23.23	22.91	22.69	22.53	22.49	22.47	22.46	22.43	22.42	22.41	22.39	22.38	22.38
.77	26.77	24.33	23.47	22.92	22.60	22.38	22.23	22.19	22.17	22.16	22.13	22.12	22.11	22.09	22.08	22.08
.78	26.39	24.00	23.15	22.60	22.30	22.08	21.93	21.90	21.88	21.86	21.85	21.83	21.81	21.80	21.78	21.78
.79	26.02	23.67	22.83	22.29	21.99	21.78	21.64	21.60	21.58	21.57	21.54	21.53	21.52	21.50	21.49	21.49
.80	25.64	23.33	22.51	21.98	21.69	21.48	21.34	21.30	21.28	21.27	21.26	21.24	21.22	21.22	21.20	21.20
.81	25.25	23.00	22.19	21.68	21.39	21.18	21.04	21.01	20.99	20.98	20.97	20.95	20.93	20.93	20.91	20.91
.82	24.86	22.67	21.87	21.37	21.09	20.89	20.75	20.72	20.70	20.69	20.68	20.66	20.64	20.64	20.62	20.62
.83	24.47	22.33	21.56	21.06	20.79	20.59	20.46	20.43	20.41	20.40	20.38	20.37	20.36	20.35	20.34	20.34
.84	24.07	22.00	21.24	20.76	20.49	20.30	20.17	20.15	20.13	20.12	20.10	20.09	20.08	20.06	20.06	20.06
.85	23.67	21.67	20.93	20.46	20.20	20.01	19.89	19.87	19.85	19.84	19.82	19.81	19.79	19.79	19.78	19.78
.86	23.26	21.33	20.62	20.16	19.90	19.73	19.60	19.58	19.57	19.56	19.54	19.54	19.52	19.51	19.50	19.50
.87	22.84	21.00	20.31	19.86	19.61	19.44	19.32	19.31	19.29	19.28	19.26	19.25	19.24	19.24	19.22	19.22
.88	22.42	20.67	20.00	19.57	19.33	19.16	19.04	19.03	19.01	19.00	18.98	18.98	18.97	18.96	18.95	18.95
.89	21.99	20.33	19.69	19.27	19.04	18.88	18.77	18.75	18.74	18.73	18.71	18.70	18.69	18.69	18.68	18.68
.90	21.55	20.00	19.38	18.98	18.75	18.60	18.50	18.48	18.47	18.46	18.44	18.43	18.43	18.42	18.41	18.41
.91	21.11	19.67	19.07	18.69	18.47	18.32	18.22	18.21	18.20	18.19	18.17	18.17	18.17	18.16	18.15	18.15
.92	20.66	19.33	18.77	18.40	18.19	18.05	17.96	17.95	17.93	17.92	17.92	17.90	17.89	17.89	17.88	17.88
.93	20.20	19.00	18.46	18.11	17.91	17.78	17.69	17.68	17.67	17.66	17.65	17.65	17.63	17.63	17.62	17.62
.94	19.74	18.67	18.16	17.82	17.64	17.51	17.43	17.42	17.41	17.40	17.39	17.39	17.37	17.37	17.36	17.36
.95	19.29	18.33	17.86	17.54	17.36	17.24	17.17	17.16	17.15	17.14	17.13	17.13	17.12	17.12	17.11	17.11
.96	18.76	18.00	17.56	17.26	17.09	16.98	16.91	16.90	16.89	16.88	16.88	16.87	16.86	16.86	16.86	16.86
.97	18.25	17.67	17.25	16.97	16.82	16.71	16.65	16.64	16.63	16.63	16.62	16.62	16.61	16.61	16.60	16.60
.98	17.74	17.33	16.95	16.70	16.55	16.45	16.39	16.38	16.38	16.37	16.37	16.37	16.36	16.36	16.36	16.36
.99	17.22	17.00	16.66	16.42	16.28	16.19	16.14	16.13	16.13	16.12	16.12	16.12	16.11	16.11	16.12	16.11
1.00	16.67	16.67	16.36	16.16	16.02	15.94	15.89	15.88	15.88	15.88	15.87	15.87	15.87	15.87	15.87	15.87
1.01	16.11	16.33	16.07	15.87	15.76	15.68	15.64	15.63	15.63	15.63	15.63	15.63	15.62	15.62	15.62	15.62
1.02	15.53	16.00	15.78	15.60	15.50	15.43	15.40	15.39	15.39	15.39	15.39	15.39	15.38	15.38	15.38	15.38
1.03	14.93	15.67	15.48	15.33	15.24	15.18	15.15	15.15	15.15	15.15	15.15	15.15	15.15	15.15	15.15	15.15
1.04	14.31	15.33	15.19	15.06	14.98	14.94	14.92	14.92	14.91	14.92	14.92	14.92	14.92	14.92	14.92	14.92
1.05	13.66	15.00	14.91	14.79	14.73	14.69	14.67	14.67	14.67	14.67	14.67	14.68	14.68	14.68	14.68	14.68
1.06	12.98	14.67	14.62	14.53	14.48	14.45	14.44	14.44	14.44	14.44	14.44	14.44	14.45	14.46	14.46	14.46
1.07	12.27	14.33	14.33	14.27	14.23	14.21	14.20	14.21	14.21	14.21	14.21	14.21	14.22	14.22	14.22	14.22
1.08	11.51	14.00	14.05	14.01	13.98	13.97	13.97	13.98	13.98	13.98	13.98	13.99	13.99	13.99	14.00	14.00
1.09	10.71	13.67	13.76	13.75	13.74	13.74	13.75	13.76	13.75	13.75	13.75	13.76	13.76	13.77	13.77	13.78

* Values tabulated are read in per cent.

TABLE VI

TABLE FOR ESTIMATING THE LOT PERCENTAGE FOR PLANS BASED ON AVERAGE RANGE—(*Continued*)

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1.10	9.46	13.39	13.40	13.50	13.49	13.50	13.52	13.52	13.52	13.53	13.54	13.54	13.55	13.55	13.56	13.56
1.11	8.89	13.00	13.20	13.24	13.25	13.27	13.29	13.30	13.30	13.31	13.31	13.32	13.32	13.33	13.34	13.34
1.12	7.86	12.67	12.93	12.99	13.02	13.04	13.07	13.08	13.08	13.09	13.10	13.10	13.12	13.12	13.12	13.12
1.13	6.40	12.33	12.65	12.74	12.78	12.81	12.85	12.86	12.86	12.87	12.89	12.89	12.89	12.90	12.91	12.91
1.14	5.00	12.00	12.37	12.49	12.55	12.59	12.63	12.64	12.65	12.66	12.67	12.67	12.69	12.69	12.70	12.70
1.15	0.29	11.67	12.10	12.25	12.31	12.37	12.42	12.43	12.44	12.45	12.46	12.46	12.48	12.48	12.49	12.49
1.16	0.00	11.33	11.83	12.00	12.08	12.15	12.21	12.22	12.23	12.24	12.25	12.25	12.27	12.28	12.29	12.29
1.17	0.00	11.00	11.56	11.76	11.86	11.93	12.00	12.01	12.02	12.03	12.05	12.06	12.07	12.07	12.08	12.08
1.18	0.00	10.67	11.29	11.52	11.63	11.72	11.79	11.80	11.81	11.82	11.84	11.84	11.86	11.88	11.88	11.88
1.19	0.00	10.33	11.02	11.29	11.41	11.50	11.58	11.60	11.61	11.62	11.64	11.65	11.66	11.68	11.69	11.69
1.20	0.00	10.00	10.74	11.03	11.19	11.28	11.36	11.40	11.43	11.42	11.43	11.45	11.47	11.47	11.49	11.49
1.21	0.00	9.67	10.50	10.83	10.97	11.08	11.16	11.20	11.21	11.22	11.25	11.26	11.27	11.29	11.30	11.30
1.22	0.00	9.33	10.23	10.59	10.76	10.88	10.98	11.00	11.02	11.03	11.04	11.06	11.08	11.09	11.10	11.10
1.23	0.00	9.00	9.97	10.36	10.54	10.67	10.78	10.80	10.82	10.84	10.85	10.87	10.89	10.90	10.91	10.91
1.24	0.00	8.67	9.72	10.13	10.33	10.47	10.58	10.61	10.63	10.64	10.67	10.68	10.70	10.71	10.73	10.73
1.25	0.00	8.33	9.46	9.91	10.12	10.27	10.39	10.42	10.44	10.46	10.47	10.49	10.51	10.52	10.54	10.54
1.26	0.00	8.00	9.21	9.69	9.92	10.08	10.20	10.24	10.26	10.27	10.30	10.31	10.33	10.34	10.36	10.36
1.27	0.00	7.67	8.96	9.47	9.71	9.88	10.01	10.05	10.07	10.09	10.11	10.13	10.15	10.17	10.18	10.18
1.28	0.00	7.33	8.71	9.25	9.51	9.69	9.83	9.87	9.89	9.90	9.93	9.95	9.97	9.99	10.00	10.00
1.29	0.00	7.00	8.46	9.04	9.31	9.50	9.64	9.68	9.71	9.72	9.75	9.77	9.79	9.81	9.83	9.83
1.30	0.00	6.67	8.21	8.83	9.11	9.31	9.47	9.51	9.53	9.55	9.58	9.59	9.62	9.64	9.65	9.65
1.31	0.00	6.33	7.97	8.62	8.92	9.13	9.29	9.33	9.35	9.37	9.40	9.42	9.45	9.47	9.48	9.48
1.32	0.00	6.00	7.73	8.41	8.73	8.95	9.11	9.15	9.18	9.20	9.23	9.25	9.28	9.30	9.31	9.31
1.33	0.00	5.67	7.49	8.20	8.54	8.77	8.94	8.98	9.01	9.03	9.06	9.08	9.11	9.13	9.14	9.15
1.34	0.00	5.33	7.25	8.00	8.35	8.59	8.77	8.81	8.84	8.86	8.89	8.91	8.94	8.96	8.98	8.98
1.35	0.00	5.00	7.02	7.80	8.16	8.42	8.60	8.64	8.67	8.69	8.73	8.75	8.78	8.80	8.82	8.82
1.36	0.00	4.67	6.79	7.60	7.98	8.24	8.43	8.48	8.51	8.53	8.56	8.59	8.62	8.64	8.66	8.66
1.37	0.00	4.33	6.56	7.40	7.80	8.07	8.27	8.31	8.34	8.37	8.40	8.43	8.46	8.48	8.50	8.50
1.38	0.00	4.00	6.33	7.21	7.62	7.90	8.11	8.15	8.18	8.21	8.25	8.26	8.30	8.32	8.34	8.35
1.39	0.00	3.67	6.10	7.02	7.45	7.73	7.95	7.99	8.02	8.05	8.09	8.11	8.14	8.17	8.19	8.19
1.40	0.00	3.33	5.88	6.83	7.27	7.57	7.79	7.84	7.88	7.90	7.93	7.96	8.00	8.02	8.03	8.04
1.41	0.00	3.00	5.66	6.65	7.10	7.41	7.63	7.68	7.72	7.74	7.78	7.81	7.84	7.87	7.88	7.88
1.42	0.00	2.67	5.44	6.46	6.93	7.25	7.48	7.53	7.56	7.59	7.63	7.66	7.70	7.72	7.74	7.74
1.43	0.00	2.33	5.23	6.28	6.76	7.09	7.33	7.38	7.41	7.44	7.48	7.51	7.55	7.57	7.59	7.60
1.44	0.00	2.00	5.01	6.10	6.60	6.93	7.18	7.24	7.28	7.30	7.34	7.37	7.41	7.43	7.45	7.46
1.45	0.00	1.67	4.81	5.93	6.44	6.78	7.03	7.09	7.13	7.16	7.20	7.23	7.27	7.29	7.31	7.31
1.46	0.00	1.33	4.60	5.75	6.28	6.63	6.89	6.95	6.99	7.02	7.06	7.09	7.13	7.15	7.17	7.18
1.47	0.00	1.00	4.39	5.58	6.12	6.48	6.74	6.80	6.85	6.87	6.92	6.95	6.99	7.01	7.03	7.04
1.48	0.00	0.67	4.19	5.41	5.96	6.34	6.60	6.66	6.72	6.73	6.78	6.81	6.85	6.87	6.89	6.90
1.49	0.00	0.33	3.99	5.24	5.81	6.19	6.47	6.53	6.57	6.60	6.64	6.67	6.72	6.74	6.76	6.77

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1.50	0.00	0.00	3.80	5.09	5.66	6.05	6.33	6.39	6.43	6.46	6.52	6.54	6.58	6.61	6.63	6.64
1.51	0.00	0.00	3.61	4.93	5.51	5.91	6.20	6.25	6.30	6.33	6.38	6.41	6.46	6.48	6.50	6.51
1.52	0.00	0.00	3.42	4.76	5.37	5.77	6.06	6.12	6.17	6.20	6.25	6.28	6.30	6.35	6.37	6.38
1.53	0.00	0.00	3.23	4.60	5.22	5.64	5.93	5.99	6.04	6.07	6.12	6.15	6.20	6.22	6.25	6.26
1.54	0.00	0.00	3.05	4.45	5.08	5.50	5.80	5.86	5.91	5.95	6.00	6.03	6.07	6.10	6.12	6.14
1.55	0.00	0.00	2.87	4.30	4.94	5.37	5.68	5.74	5.79	5.82	5.87	5.90	5.95	5.98	6.00	6.01
1.56	0.00	0.00	2.69	4.15	4.81	5.24	5.55	5.62	5.67	5.70	5.75	5.78	5.83	5.86	5.88	5.89
1.57	0.00	0.00	2.53	4.01	4.67	5.11	5.43	5.50	5.55	5.58	5.63	5.66	5.71	5.74	5.77	5.79
1.58	0.00	0.00	2.35	3.86	4.54	4.99	5.31	5.38	5.43	5.46	5.52	5.55	5.59	5.62	5.65	5.66
1.59	0.00	0.00	2.19	3.72	4.41	4.86	5.19	5.26	5.31	5.34	5.40	5.43	5.48	5.51	5.53	5.55
1.60	0.00	0.00	2.03	3.58	4.28	4.74	5.08	5.14	5.19	5.23	5.29	5.32	5.36	5.39	5.42	5.43
1.61	0.00	0.00	1.87	3.45	4.16	4.62	4.96	5.03	5.08	5.12	5.17	5.20	5.25	5.28	5.31	5.32
1.62	0.00	0.00	1.72	3.31	4.03	4.51	4.85	4.92	4.97	5.01	5.06	5.09	5.14	5.17	5.20	5.20
1.63	0.00	0.00	1.57	3.18	3.91	4.39	4.74	4.81	4.86	4.90	4.96	4.99	5.04	5.07	5.10	5.12
1.64	0.00	0.00	1.42	3.05	3.79	4.28	4.63	4.70	4.75	4.79	4.85	4.88	4.93	4.96	4.99	5.00
1.65	0.00	0.00	1.28	2.93	3.68	4.17	4.52	4.59	4.64	4.68	4.74	4.77	4.83	4.86	4.89	4.91
1.66	0.00	0.00	1.15	2.81	3.56	4.06	4.42	4.49	4.54	4.58	4.64	4.67	4.72	4.75	4.79	4.81
1.67	0.00	0.00	1.02	2.69	3.45	3.95	4.31	4.39	4.44	4.48	4.54	4.57	4.62	4.65	4.69	4.71
1.68	0.00	0.00	0.89	2.57	3.34	3.85	4.21	4.29	4.34	4.38	4.44	4.47	4.53	4.56	4.59	4.61
1.69	0.00	0.00	0.77	2.46	3.23	3.74	4.10	4.19	4.24	4.28	4.34	4.37	4.43	4.46	4.49	4.51
1.70	0.00	0.00	0.66	2.35	3.13	3.64	4.00	4.09	4.14	4.18	4.24	4.28	4.33	4.36	4.40	4.42
1.71	0.00	0.00	0.55	2.24	3.02	3.54	3.92	3.99	4.05	4.09	4.15	4.18	4.24	4.27	4.30	4.32
1.72	0.00	0.00	0.45	2.13	2.92	3.45	3.82	3.90	3.95	3.99	4.06	4.09	4.15	4.18	4.21	4.23
1.73	0.00	0.00	0.36	2.03	2.82	3.35	3.73	3.81	3.86	3.90	3.96	4.00	4.06	4.09	4.12	4.14
1.74	0.00	0.00	0.27	1.93	2.73	3.26	3.63	3.72	3.77	3.81	3.87	3.91	3.97	4.00	4.03	4.05
1.75	0.00	0.00	0.19	1.83	2.63	3.16	3.54	3.63	3.68	3.72	3.79	3.82	3.88	3.91	3.94	3.96
1.76	0.00	0.00	0.12	1.73	2.54	3.07	3.45	3.54	3.59	3.63	3.70	3.74	3.79	3.82	3.86	3.88
1.77	0.00	0.00	0.06	1.64	2.45	2.99	3.37	3.45	3.51	3.55	3.61	3.65	3.71	3.74	3.77	3.79
1.78	0.00	0.00	0.02	1.55	2.36	2.90	3.28	3.37	3.43	3.47	3.53	3.57	3.62	3.65	3.69	3.72
1.79	0.00	0.00	0.00	1.46	2.27	2.81	3.20	3.28	3.34	3.38	3.45	3.49	3.54	3.57	3.61	3.63
1.80	0.00	0.00	0.00	1.38	2.19	2.73	3.12	3.20	3.26	3.30	3.37	3.41	3.46	3.49	3.53	3.55
1.81	0.00	0.00	0.00	1.29	2.10	2.64	3.03	3.12	3.18	3.22	3.29	3.33	3.38	3.42	3.45	3.47
1.82	0.00	0.00	0.00	1.21	2.02	2.57	2.96	3.05	3.11	3.15	3.22	3.25	3.31	3.34	3.37	3.39
1.83	0.00	0.00	0.00	1.14	1.94	2.49	2.88	2.97	3.03	3.07	3.13	3.17	3.23	3.26	3.30	3.32
1.84	0.00	0.00	0.00	1.06	1.87	2.42	2.80	2.89	2.95	2.99	3.06	3.10	3.16	3.19	3.22	3.24
1.85	0.00	0.00	0.00	0.99	1.79	2.34	2.73	2.82	2.88	2.92	2.99	3.03	3.08	3.12	3.15	3.17
1.86	0.00	0.00	0.00	0.92	1.72	2.27	2.66	2.75	2.81	2.85	2.92	2.95	3.01	3.05	3.08	3.10
1.87	0.00	0.00	0.00	0.86	1.65	2.20	2.59	2.68	2.74	2.78	2.84	2.88	2.94	2.97	3.01	3.03
1.88	0.00	0.00	0.00	0.79	1.58	2.13	2.52	2.61	2.67	2.71	2.77	2.81	2.87	2.90	2.94	2.96
1.89	0.00	0.00	0.00	0.73	1.51	2.06	2.45	2.54	2.60	2.64	2.71	2.75	2.81	2.84	2.87	2.89

TABLE VI

TABLE FOR ESTIMATING THE LOT PERCENTAGE FOR PLANS BASED ON AVERAGE RANGE—(*Continued*)

Q_U'' or Q_L''	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1.90	0.00	0.00	0.00	0.67	1.45	1.99	2.38	2.47	2.53	2.57	2.64	2.68	2.74	2.77	2.81	2.83
1.91	0.00	0.00	0.00	0.62	1.38	1.93	2.32	2.41	2.47	2.51	2.58	2.61	2.67	2.70	2.74	2.76
1.92	0.00	0.00	0.00	0.56	1.32	1.86	2.25	2.34	2.41	2.45	2.51	2.55	2.61	2.64	2.68	2.70
1.93	0.00	0.00	0.00	0.51	1.26	1.80	2.19	2.28	2.34	2.38	2.45	2.49	2.55	2.58	2.61	2.63
1.94	0.00	0.00	0.00	0.46	1.20	1.74	2.13	2.22	2.28	2.32	2.39	2.43	2.49	2.52	2.55	2.57
1.95	0.00	0.00	0.00	0.42	1.15	1.68	2.07	2.16	2.22	2.26	2.33	2.37	2.43	2.46	2.49	2.51
1.96	0.00	0.00	0.00	0.37	1.09	1.62	2.01	2.10	2.16	2.20	2.27	2.31	2.37	2.40	2.43	2.45
1.97	0.00	0.00	0.00	0.33	1.04	1.57	1.95	2.04	2.10	2.14	2.21	2.25	2.31	2.34	2.38	2.40
1.98	0.00	0.00	0.00	0.30	0.99	1.51	1.90	1.99	2.05	2.09	2.15	2.19	2.25	2.28	2.32	2.34
1.99	0.00	0.00	0.00	0.26	0.94	1.46	1.84	1.93	1.99	2.03	2.10	2.14	2.20	2.23	2.26	2.28
2.00	0.00	0.00	0.00	0.23	0.89	1.41	1.79	1.88	1.94	1.98	2.05	2.08	2.14	2.17	2.21	2.23
2.01	0.00	0.00	0.00	0.20	0.84	1.36	1.74	1.83	1.89	1.93	1.99	2.03	2.09	2.12	2.16	2.18
2.02	0.00	0.00	0.00	0.17	0.80	1.31	1.69	1.78	1.83	1.87	1.94	1.98	2.04	2.07	2.10	2.12
2.03	0.00	0.00	0.00	0.14	0.75	1.26	1.64	1.73	1.78	1.82	1.89	1.93	1.99	2.02	2.05	2.07
2.04	0.00	0.00	0.00	0.12	0.71	1.21	1.59	1.68	1.73	1.77	1.84	1.88	1.94	1.97	2.00	2.02
2.05	0.00	0.00	0.00	0.10	0.67	1.17	1.54	1.63	1.69	1.73	1.79	1.83	1.89	1.92	1.95	1.97
2.06	0.00	0.00	0.00	0.08	0.63	1.12	1.49	1.58	1.64	1.68	1.74	1.78	1.84	1.87	1.91	1.93
2.07	0.00	0.00	0.00	0.06	0.60	1.08	1.45	1.54	1.59	1.63	1.70	1.74	1.79	1.83	1.86	1.88
2.08	0.00	0.00	0.00	0.05	0.56	1.04	1.40	1.49	1.55	1.59	1.65	1.69	1.75	1.78	1.81	1.83
2.09	0.00	0.00	0.00	0.03	0.53	1.00	1.36	1.45	1.50	1.54	1.61	1.64	1.70	1.73	1.77	1.79
2.10	0.00	0.00	0.00	0.02	0.47	0.96	1.32	1.42	1.46	1.50	1.56	1.60	1.66	1.69	1.72	1.76
2.11	0.00	0.00	0.00	0.01	0.46	0.92	1.28	1.36	1.42	1.46	1.52	1.56	1.61	1.66	1.68	1.70
2.12	0.00	0.00	0.00	0.00	0.43	0.88	1.24	1.32	1.38	1.42	1.48	1.52	1.57	1.60	1.64	1.66
2.13	0.00	0.00	0.00	0.00	0.40	0.85	1.20	1.28	1.34	1.38	1.44	1.48	1.53	1.56	1.60	1.62
2.14	0.00	0.00	0.00	0.00	0.38	0.82	1.16	1.25	1.30	1.34	1.40	1.44	1.49	1.52	1.56	1.58
2.15	0.00	0.00	0.00	0.00	0.35	0.79	1.13	1.21	1.26	1.30	1.36	1.40	1.45	1.48	1.52	1.54
2.16	0.00	0.00	0.00	0.00	0.32	0.75	1.09	1.17	1.22	1.26	1.32	1.36	1.41	1.44	1.48	1.50
2.17	0.00	0.00	0.00	0.00	0.30	0.71	1.06	1.13	1.18	1.22	1.29	1.32	1.38	1.41	1.44	1.46
2.18	0.00	0.00	0.00	0.00	0.28	0.68	1.02	1.10	1.15	1.19	1.25	1.28	1.34	1.37	1.40	1.42
2.19	0.00	0.00	0.00	0.00	0.26	0.65	0.99	1.06	1.11	1.15	1.22	1.25	1.30	1.33	1.37	1.39
2.20	0.000	0.000	0.000	0.000	0.236	0.623	0.954	1.030	1.083	1.122	1.178	1.217	1.267	1.299	1.330	1.346
2.21	0.000	0.000	0.000	0.000	0.217	0.597	0.923	0.997	1.050	1.089	1.144	1.180	1.233	1.265	1.295	1.311
2.22	0.000	0.000	0.000	0.000	0.199	0.570	0.892	0.966	1.018	1.056	1.111	1.147	1.199	1.231	1.261	1.277
2.23	0.000	0.000	0.000	0.000	0.182	0.544	0.861	0.935	0.986	1.025	1.079	1.115	1.167	1.197	1.228	1.244
2.24	0.000	0.000	0.000	0.000	0.166	0.519	0.831	0.905	0.956	0.994	1.048	1.083	1.135	1.166	1.196	1.211
2.25	0.000	0.000	0.000	0.000	0.150	0.495	0.802	0.875	0.926	0.964	1.018	1.052	1.104	1.134	1.163	1.179
2.26	0.000	0.000	0.000	0.000	0.136	0.471	0.773	0.847	0.897	0.935	0.987	1.022	1.073	1.103	1.132	1.148
2.27	0.000	0.000	0.000	0.000	0.123	0.449	0.746	0.819	0.869	0.906	0.958	0.993	1.043	1.073	1.102	1.118
2.28	0.000	0.000	0.000	0.000	0.111	0.427	0.720	0.792	0.841	0.878	0.930	0.964	1.014	1.044	1.073	1.088
2.29	0.000	0.000	0.000	0.000	0.099	0.406	0.697	0.766	0.814	0.851	0.902	0.936	0.986	1.015	1.044	1.059

Q_U'' or Q_L''	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
2.30	0.000	0.000	0.000	0.000	0.089	0.386	0.672	0.741	0.789	0.825	0.875	0.909	0.959	0.988	1.016	1.031
2.31	0.000	0.000	0.000	0.000	0.079	0.367	0.649	0.716	0.763	0.799	0.849	0.882	0.931	0.960	0.987	1.003
2.32	0.000	0.000	0.000	0.000	0.070	0.348	0.626	0.692	0.739	0.774	0.823	0.856	0.905	0.934	0.960	0.976
2.33	0.000	0.000	0.000	0.000	0.061	0.330	0.603	0.668	0.715	0.750	0.798	0.831	0.879	0.908	0.935	0.950
2.34	0.000	0.000	0.000	0.000	0.054	0.313	0.581	0.645	0.691	0.726	0.774	0.807	0.854	0.882	0.909	0.924
2.35	0.000	0.000	0.000	0.000	0.047	0.296	0.559	0.623	0.669	0.703	0.750	0.782	0.829	0.857	0.884	0.899
2.36	0.000	0.000	0.000	0.000	0.040	0.280	0.538	0.602	0.646	0.680	0.728	0.759	0.806	0.833	0.860	0.875
2.37	0.000	0.000	0.000	0.000	0.035	0.265	0.518	0.580	0.624	0.658	0.705	0.736	0.782	0.809	0.836	0.850
2.38	0.000	0.000	0.000	0.000	0.029	0.250	0.498	0.560	0.604	0.637	0.683	0.714	0.759	0.787	0.813	0.827
2.39	0.000	0.000	0.000	0.000	0.025	0.236	0.479	0.541	0.584	0.616	0.662	0.693	0.737	0.764	0.791	0.804
2.40	0.000	0.000	0.000	0.000	0.022	0.223	0.461	0.521	0.564	0.596	0.641	0.671	0.715	0.742	0.769	0.782
2.41	0.000	0.000	0.000	0.000	0.017	0.210	0.443	0.503	0.545	0.577	0.621	0.651	0.695	0.721	0.747	0.760
2.42	0.000	0.000	0.000	0.000	0.014	0.198	0.426	0.485	0.526	0.557	0.601	0.631	0.674	0.701	0.726	0.739
2.43	0.000	0.000	0.000	0.000	0.011	0.186	0.410	0.467	0.508	0.539	0.582	0.611	0.654	0.680	0.705	0.718
2.44	0.000	0.000	0.000	0.000	0.009	0.175	0.393	0.450	0.491	0.521	0.564	0.593	0.635	0.660	0.685	0.698
2.45	0.000	0.000	0.000	0.000	0.007	0.165	0.378	0.434	0.473	0.503	0.545	0.573	0.616	0.641	0.666	0.678
2.46	0.000	0.000	0.000	0.000	0.005	0.154	0.362	0.417	0.456	0.486	0.528	0.556	0.597	0.622	0.646	0.659
2.47	0.000	0.000	0.000	0.000	0.004	0.145	0.348	0.403	0.441	0.470	0.511	0.539	0.579	0.604	0.627	0.640
2.48	0.000	0.000	0.000	0.000	0.003	0.136	0.333	0.387	0.425	0.454	0.494	0.522	0.562	0.586	0.609	0.622
2.49	0.000	0.000	0.000	0.000	0.002	0.127	0.321	0.372	0.409	0.438	0.478	0.504	0.546	0.569	0.593	0.605
2.50	0.000	0.000	0.000	0.000	0.001	0.118	0.307	0.358	0.395	0.423	0.463	0.489	0.528	0.552	0.575	0.587
2.51	0.000	0.000	0.000	0.000	0.001	0.111	0.294	0.345	0.381	0.409	0.447	0.473	0.512	0.536	0.558	0.570
2.52	0.000	0.000	0.000	0.000	0.000	0.103	0.282	0.332	0.367	0.394	0.432	0.458	0.497	0.519	0.542	0.553
2.53	0.000	0.000	0.000	0.000	0.000	0.096	0.270	0.319	0.354	0.380	0.418	0.444	0.481	0.503	0.526	0.537
2.54	0.000	0.000	0.000	0.000	0.000	0.089	0.258	0.306	0.340	0.367	0.404	0.429	0.466	0.488	0.510	0.521
2.55	0.000	0.000	0.000	0.000	0.000	0.083	0.247	0.294	0.328	0.354	0.390	0.415	0.451	0.473	0.495	0.506
2.56	0.000	0.000	0.000	0.000	0.000	0.077	0.237	0.283	0.316	0.341	0.377	0.401	0.437	0.459	0.480	0.491
2.57	0.000	0.000	0.000	0.000	0.000	0.071	0.227	0.272	0.304	0.329	0.364	0.388	0.424	0.445	0.466	0.477
2.58	0.000	0.000	0.000	0.000	0.000	0.066	0.217	0.261	0.292	0.317	0.352	0.376	0.411	0.432	0.452	0.463
2.59	0.000	0.000	0.000	0.000	0.000	0.061	0.207	0.251	0.282	0.305	0.340	0.363	0.397	0.418	0.439	0.449
2.60	0.000	0.000	0.000	0.000	0.000	0.056	0.198	0.240	0.271	0.294	0.328	0.351	0.385	0.405	0.426	0.436
2.61	0.000	0.000	0.000	0.000	0.000	0.052	0.189	0.231	0.260	0.283	0.317	0.339	0.372	0.393	0.413	0.423
2.62	0.000	0.000	0.000	0.000	0.000	0.048	0.181	0.222	0.250	0.273	0.306	0.327	0.360	0.381	0.400	0.410
2.63	0.000	0.000	0.000	0.000	0.000	0.044	0.173	0.213	0.241	0.263	0.295	0.316	0.349	0.368	0.388	0.398
2.64	0.000	0.000	0.000	0.000	0.000	0.040	0.164	0.204	0.232	0.253	0.285	0.306	0.338	0.357	0.376	0.386
2.65	0.000	0.000	0.000	0.000	0.000	0.037	0.157	0.195	0.223	0.244	0.274	0.295	0.327	0.346	0.365	0.375
2.66	0.000	0.000	0.000	0.000	0.000	0.034	0.149	0.188	0.213	0.234	0.265	0.285	0.316	0.335	0.353	0.363
2.67	0.000	0.000	0.000	0.000	0.000	0.031	0.142	0.179	0.205	0.225	0.255	0.275	0.305	0.324	0.342	0.352
2.68	0.000	0.000	0.000	0.000	0.000	0.028	0.136	0.171	0.197	0.217	0.246	0.266	0.296	0.314	0.332	0.342
2.69	0.000	0.000	0.000	0.000	0.000	0.025	0.129	0.164	0.189	0.209	0.238	0.257	0.286	0.304	0.321	0.331

TABLE VI

TABLE FOR ESTIMATING THE LOT PERCENTAGE FOR PLANS BASED ON AVERAGE RANGE—(*Continued*)

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
2.70	0.000	0.000	0.000	0.000	0.000	0.023	0.123	0.156	0.182	0.201	0.228	0.248	0.277	0.295	0.311	0.321
2.71	0.000	0.000	0.000	0.000	0.000	0.021	0.117	0.150	0.176	0.193	0.220	0.239	0.267	0.285	0.302	0.311
2.72	0.000	0.000	0.000	0.000	0.000	0.019	0.111	0.143	0.167	0.185	0.212	0.231	0.259	0.275	0.292	0.301
2.73	0.000	0.000	0.000	0.000	0.000	0.017	0.106	0.137	0.160	0.178	0.205	0.222	0.250	0.266	0.283	0.292
2.74	0.000	0.000	0.000	0.000	0.000	0.015	0.101	0.132	0.153	0.171	0.197	0.213	0.242	0.258	0.274	0.282
2.75	0.000	0.000	0.000	0.000	0.000	0.014	0.096	0.125	0.147	0.164	0.189	0.207	0.233	0.248	0.266	0.274
2.76	0.000	0.000	0.000	0.000	0.000	0.012	0.091	0.120	0.141	0.158	0.182	0.200	0.225	0.241	0.257	0.265
2.77	0.000	0.000	0.000	0.000	0.000	0.011	0.086	0.114	0.135	0.152	0.175	0.192	0.217	0.232	0.249	0.257
2.78	0.000	0.000	0.000	0.000	0.000	0.010	0.081	0.109	0.130	0.146	0.169	0.185	0.210	0.226	0.241	0.249
2.79	0.000	0.000	0.000	0.000	0.000	0.008	0.077	0.103	0.124	0.140	0.163	0.179	0.203	0.218	0.233	0.241
2.80	0.000	0.000	0.000	0.000	0.000	0.007	0.074	0.099	0.118	0.134	0.156	0.172	0.196	0.210	0.225	0.233
2.81	0.000	0.000	0.000	0.000	0.000	0.007	0.070	0.094	0.113	0.129	0.150	0.165	0.189	0.204	0.218	0.226
2.82	0.000	0.000	0.000	0.000	0.000	0.006	0.066	0.090	0.109	0.123	0.144	0.159	0.183	0.196	0.211	0.219
2.83	0.000	0.000	0.000	0.000	0.000	0.005	0.062	0.085	0.103	0.118	0.139	0.154	0.176	0.190	0.204	0.212
2.84	0.000	0.000	0.000	0.000	0.000	0.004	0.059	0.082	0.099	0.113	0.134	0.148	0.170	0.184	0.197	0.205
2.85	0.000	0.000	0.000	0.000	0.000	0.004	0.055	0.078	0.095	0.109	0.128	0.143	0.164	0.178	0.191	0.198
2.86	0.000	0.000	0.000	0.000	0.000	0.003	0.053	0.074	0.091	0.104	0.124	0.137	0.159	0.172	0.185	0.192
2.87	0.000	0.000	0.000	0.000	0.000	0.003	0.050	0.070	0.087	0.100	0.119	0.132	0.152	0.166	0.179	0.185
2.88	0.000	0.000	0.000	0.000	0.000	0.002	0.047	0.067	0.082	0.095	0.114	0.127	0.147	0.160	0.173	0.179
2.89	0.000	0.000	0.000	0.000	0.000	0.002	0.044	0.064	0.079	0.091	0.109	0.122	0.142	0.155	0.167	0.173
2.90	0.000	0.000	0.000	0.000	0.000	0.002	0.042	0.061	0.075	0.088	0.105	0.117	0.138	0.149	0.161	0.168
2.91	0.000	0.000	0.000	0.000	0.000	0.001	0.039	0.057	0.072	0.084	0.101	0.113	0.132	0.145	0.156	0.162
2.92	0.000	0.000	0.000	0.000	0.000	0.001	0.037	0.055	0.069	0.080	0.097	0.107	0.127	0.140	0.151	0.157
2.93	0.000	0.000	0.000	0.000	0.000	0.001	0.035	0.052	0.066	0.077	0.093	0.104	0.123	0.134	0.146	0.151
2.94	0.000	0.000	0.000	0.000	0.000	0.001	0.033	0.049	0.062	0.073	0.089	0.100	0.118	0.129	0.141	0.146
2.95	0.000	0.000	0.000	0.000	0.000	0.001	0.031	0.047	0.059	0.070	0.086	0.096	0.114	0.125	0.136	0.142
2.96	0.000	0.000	0.000	0.000	0.000	0.001	0.029	0.044	0.056	0.067	0.082	0.092	0.110	0.121	0.132	0.137
2.97	0.000	0.000	0.000	0.000	0.000	0.000	0.027	0.042	0.054	0.064	0.079	0.089	0.105	0.116	0.127	0.132
2.98	0.000	0.000	0.000	0.000	0.000	0.000	0.025	0.039	0.051	0.061	0.075	0.085	0.101	0.112	0.123	0.128
2.99	0.000	0.000	0.000	0.000	0.000	0.000	0.024	0.038	0.049	0.058	0.072	0.082	0.098	0.108	0.119	0.124
3.00	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.036	0.047	0.056	0.069	0.078	0.094	0.105	0.115	0.120
3.01	0.000	0.000	0.000	0.000	0.000	0.000	0.022	0.034	0.044	0.053	0.066	0.075	0.091	0.101	0.111	0.116
3.02	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.032	0.042	0.050	0.063	0.072	0.087	0.097	0.107	0.112
3.03	0.000	0.000	0.000	0.000	0.000	0.000	0.019	0.030	0.040	0.048	0.061	0.069	0.084	0.094	0.103	0.108
3.04	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.028	0.038	0.045	0.058	0.066	0.081	0.090	0.099	0.104
3.05	0.000	0.000	0.000	0.000	0.000	0.000	0.016	0.027	0.036	0.043	0.056	0.064	0.078	0.086	0.096	0.101
3.06	0.000	0.000	0.000	0.000	0.000	0.000	0.015	0.025	0.034	0.041	0.053	0.061	0.075	0.083	0.092	0.097
3.07	0.000	0.000	0.000	0.000	0.000	0.000	0.014	0.024	0.032	0.039	0.051	0.059	0.072	0.080	0.089	0.094
3.08	0.000	0.000	0.000	0.000	0.000	0.000	0.013	0.022	0.030	0.037	0.049	0.056	0.069	0.077	0.086	0.091
3.09	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.021	0.029	0.036	0.046	0.054	0.067	0.075	0.083	0.088

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
3.10	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.020	0.027	0.034	0.044	0.051	0.064	0.072	0.080	0.085
3.11	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.019	0.026	0.032	0.042	0.050	0.061	0.069	0.077	0.082
3.12	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.018	0.025	0.031	0.041	0.048	0.060	0.067	0.076	0.079
3.13	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.017	0.024	0.029	0.039	0.046	0.057	0.064	0.072	0.075
3.14	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.015	0.022	0.028	0.037	0.044	0.055	0.062	0.069	0.073
3.15	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.014	0.021	0.026	0.036	0.043	0.053	0.060	0.067	0.070
3.16	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.014	0.020	0.025	0.034	0.040	0.051	0.057	0.064	0.067
3.17	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.013	0.019	0.024	0.033	0.038	0.049	0.056	0.062	0.065
3.18	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.012	0.017	0.022	0.031	0.036	0.046	0.053	0.060	0.063
3.19	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.012	0.017	0.021	0.030	0.034	0.044	0.052	0.057	0.060
3.20	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.011	0.016	0.020	0.028	0.033	0.043	0.049	0.055	0.058
3.21	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.010	0.015	0.019	0.027	0.032	0.041	0.047	0.053	0.056
3.22	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.009	0.014	0.018	0.025	0.031	0.040	0.045	0.051	0.054
3.23	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.009	0.013	0.017	0.024	0.029	0.037	0.043	0.049	0.052
3.24	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.008	0.013	0.016	0.023	0.028	0.037	0.042	0.047	0.050
3.25	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.008	0.012	0.015	0.022	0.027	0.035	0.040	0.046	0.048
3.26	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.007	0.011	0.015	0.021	0.025	0.033	0.039	0.044	0.047
3.27	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.007	0.011	0.014	0.021	0.024	0.032	0.037	0.042	0.045
3.28	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.006	0.010	0.013	0.019	0.023	0.031	0.036	0.040	0.043
3.29	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.006	0.009	0.012	0.018	0.023	0.029	0.034	0.039	0.042
3.30	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.005	0.008	0.012	0.017	0.021	0.028	0.033	0.037	0.040
3.31	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.005	0.008	0.011	0.017	0.021	0.027	0.032	0.036	0.039
3.32	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.004	0.007	0.010	0.016	0.020	0.026	0.030	0.034	0.037
3.33	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.004	0.007	0.010	0.015	0.019	0.025	0.029	0.033	0.036
3.34	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.004	0.007	0.009	0.014	0.018	0.024	0.028	0.032	0.035
3.35	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.004	0.006	0.009	0.014	0.017	0.023	0.027	0.031	0.033
3.36	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.004	0.006	0.008	0.013	0.016	0.022	0.026	0.030	0.032
3.37	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.004	0.006	0.008	0.012	0.015	0.021	0.024	0.028	0.031
3.38	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.005	0.007	0.012	0.014	0.019	0.024	0.027	0.030
3.39	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.005	0.007	0.011	0.014	0.019	0.022	0.027	0.029
3.40	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.004	0.007	0.010	0.013	0.018	0.021	0.026	0.028
3.41	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.006	0.010	0.012	0.018	0.021	0.025	0.027
3.42	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.006	0.009	0.012	0.017	0.020	0.024	0.026
3.43	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.005	0.009	0.011	0.016	0.019	0.023	0.025
3.44	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.005	0.008	0.011	0.015	0.018	0.022	0.024
3.45	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.004	0.005	0.008	0.011	0.014	0.017	0.021	0.023
3.46	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.005	0.008	0.010	0.014	0.017	0.020	0.022
3.47	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.004	0.007	0.010	0.014	0.016	0.019	0.021
3.48	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.004	0.007	0.009	0.013	0.015	0.018	0.020
3.49	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.004	0.006	0.009	0.012	0.015	0.018	0.020

TABLE VI

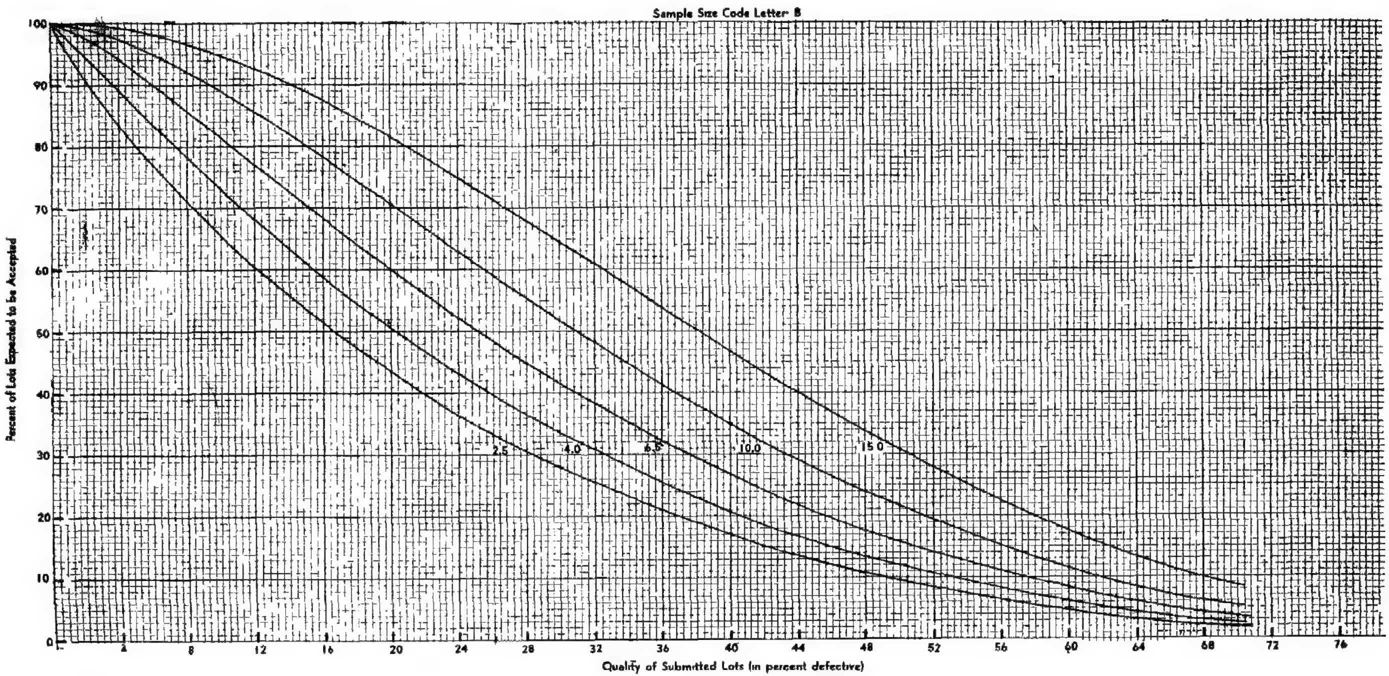
TABLE FOR ESTIMATING THE LOT PERCENTAGE FOR PLANS BASED ON AVERAGE RANGE—(*Continued*)

Q_U or Q_L	Sample Size Code Letter															
	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
3.50	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.006	0.008	0.012	0.014	0.017	0.019
3.51	0.000	0.000	0.000	0.0000	0.000	0.000	0.000	0.001	0.002	0.003	0.006	0.008	0.011	0.014	0.016	0.018
3.52	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.006	0.007	0.010	0.013	0.016	0.017
3.53	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.005	0.007	0.010	0.013	0.015	0.016
3.54	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.004	0.007	0.010	0.0 2	0.014	0.015
3.55	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.004	0.006	0.009	0.012	0.014	0.015
3.56	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.004	0.006	0.009	0.011	0.013	0.014
3.57	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.004	0.005	0.008	0.011	0.012	0.013
3.58	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.008	0.010	0.012	0.013
3.59	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.008	0.010	0.011	0.012
3.60	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.007	0.009	0.011	0.012
3.61	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.007	0.009	0.011	0.011
3.62	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.007	0.009	0.010	0.011
3.63	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.003	0.004	0.006	0.008	0.010	0.010
3.64	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.003	0.003	0.006	0.008	0.009	0.010
3.65	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.003	0.003	0.006	0.008	0.009	0.010
3.66	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.003	0.005	0.007	0.009	0.009
3.67	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.003	0.005	0.007	0.008	0.009
3.68	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.005	0.006	0.008	0.008
3.69	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.002	0.004	0.006	0.008	0.008
3.70	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.002	0.004	0.006	0.007	0.008
3.71	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.004	0.006	0.007	0.007
3.72	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.004	0.006	0.007	0.007
3.73	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.004	0.006	0.007	0.007
3.74	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.004	0.005	0.006	0.007
3.75	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.006	0.006
3.76	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.006	0.006
3.77	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.005	0.006	0.006
3.78	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.004	0.005	0.006
3.79	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.003	0.005	0.005
3.80	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.003	0.005	0.005
3.81	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.003	0.005	0.005
3.82	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.003	0.005	0.005
3.83	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.003	0.003	0.004	0.004
3.84	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.004	0.004
3.85	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.002	0.004	0.004
3.86	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.004	0.004
3.87	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.004	0.004
3.88	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.002	0.004	0.004
3.89	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.003
3.90	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.003

Fig. 1. Sampling Plans for Sample Size Code Letter: B

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)



The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution

Note Figures on curves are Acceptance Quality Levels for norm 1 inspection

Fig 2 Sampling Plans for Sample Size Code Letter C—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)

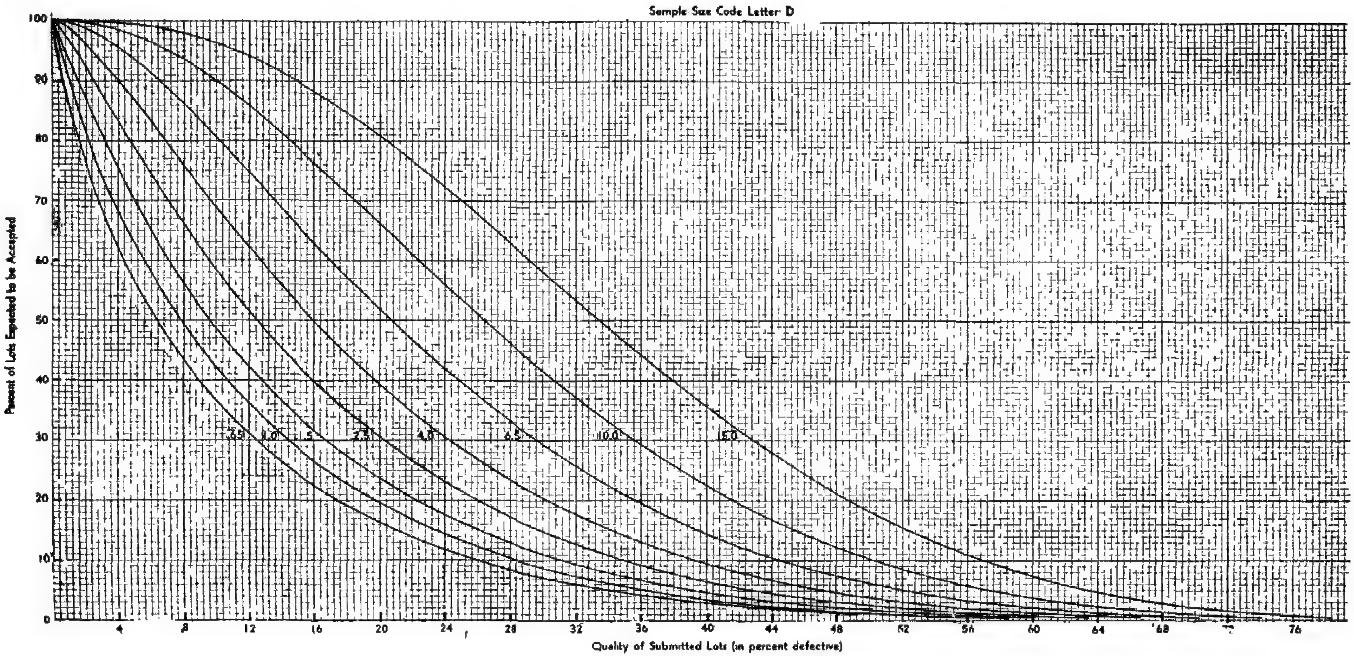
The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.

Note: Figures on curves are Acceptance Quality Levels for normal inspection.

Fig. 3 Sampling Plans for Sample Size Code Letter: D—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)

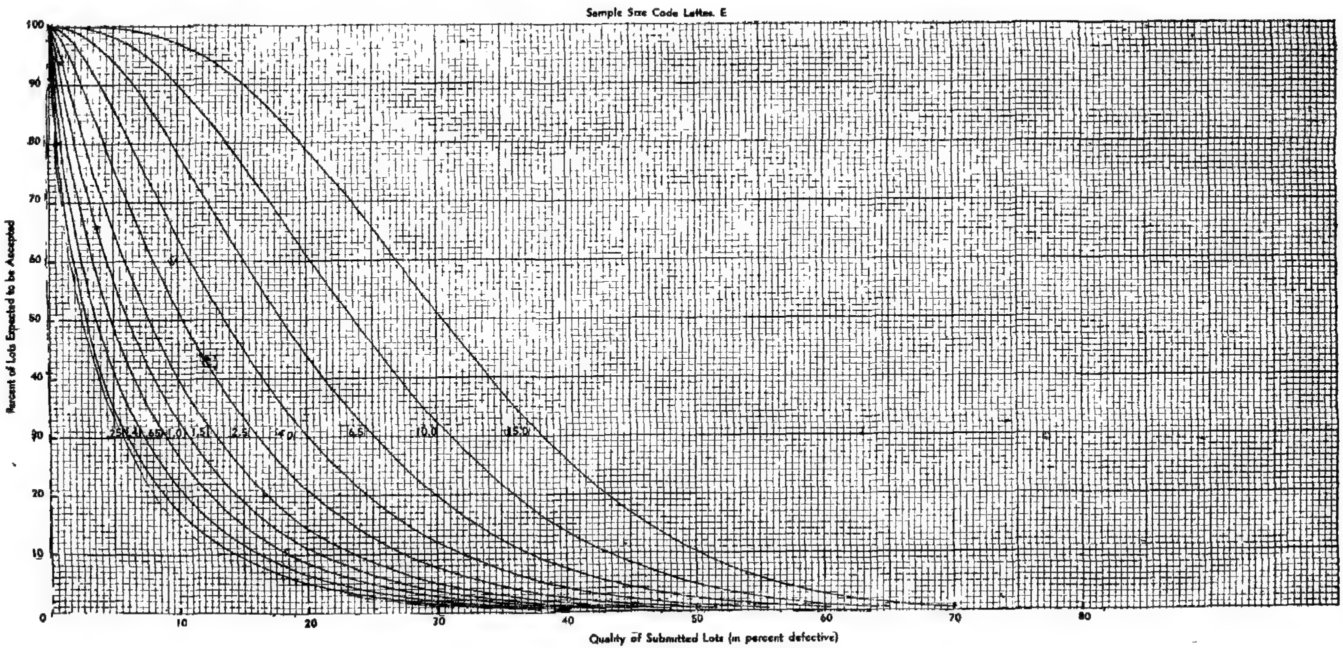


The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.
No Figur on curves are Acceptance C ality Levels for normal u st ct n

FIG. 4 Sampling Plans for Sample Size Code Letter: E—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)



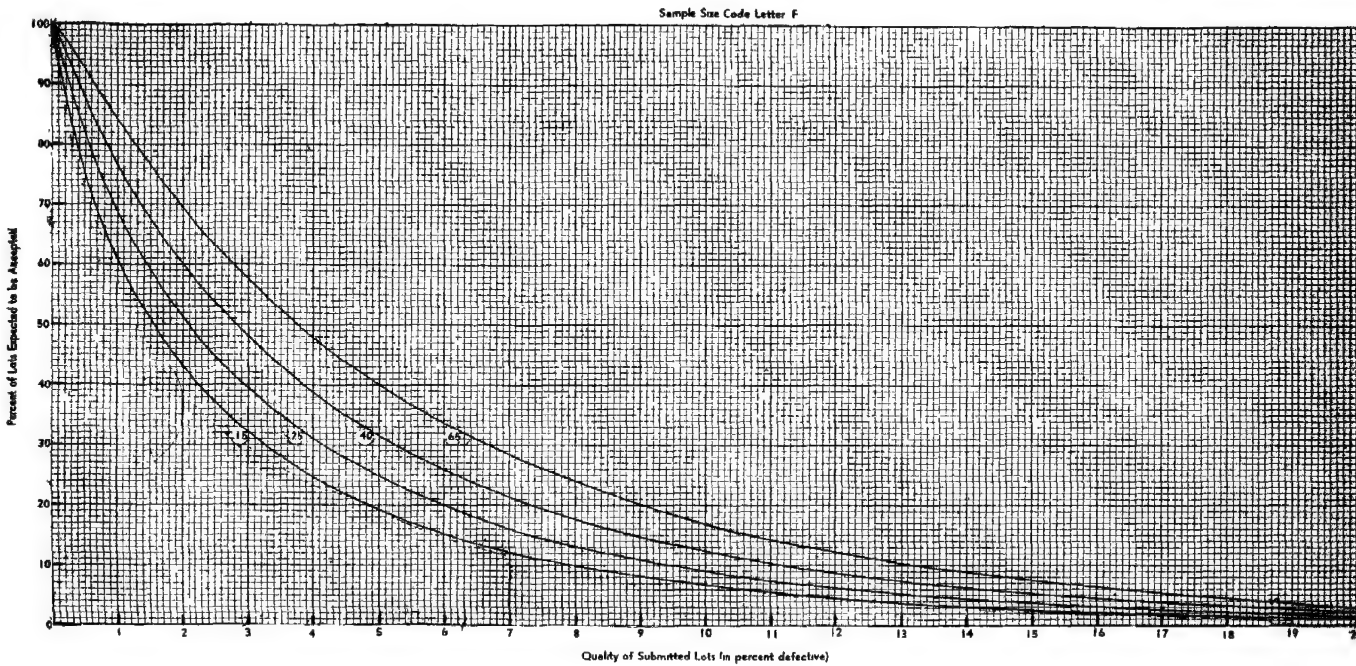
The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.

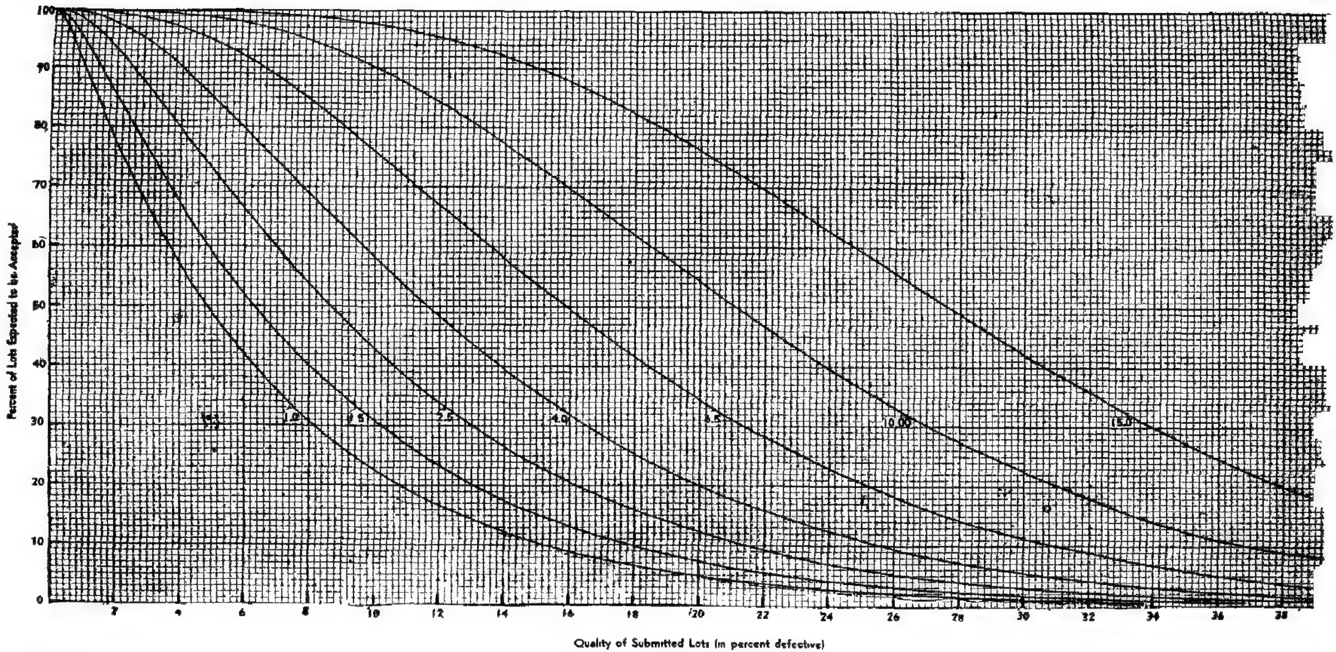
Note: Figures on curves are Acceptance Quality Levels for normal inspection.

Fig. 5. Sampling Plans for Sample Size Code Letter: F—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)



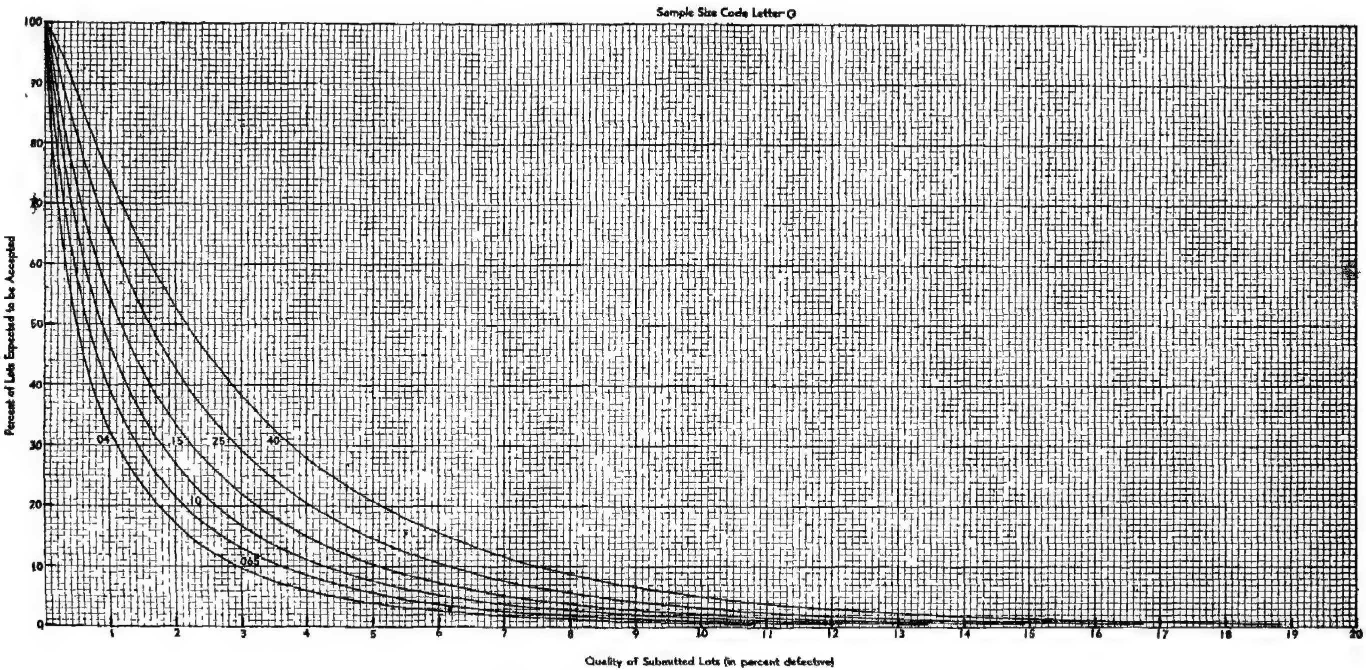


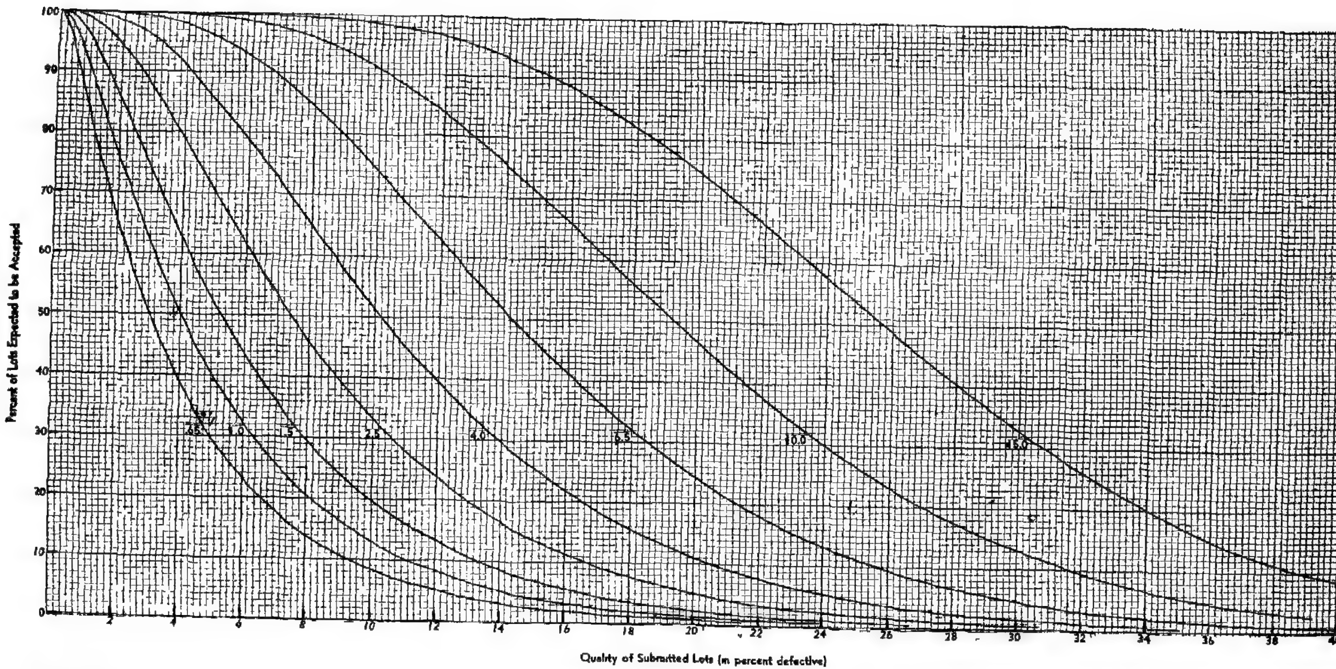
The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.
Note: Figures on curves are Acceptance Quality Levels for normal inspection.

Fig. 6 Sampling Plans for Sample Size Code Letter: G—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)



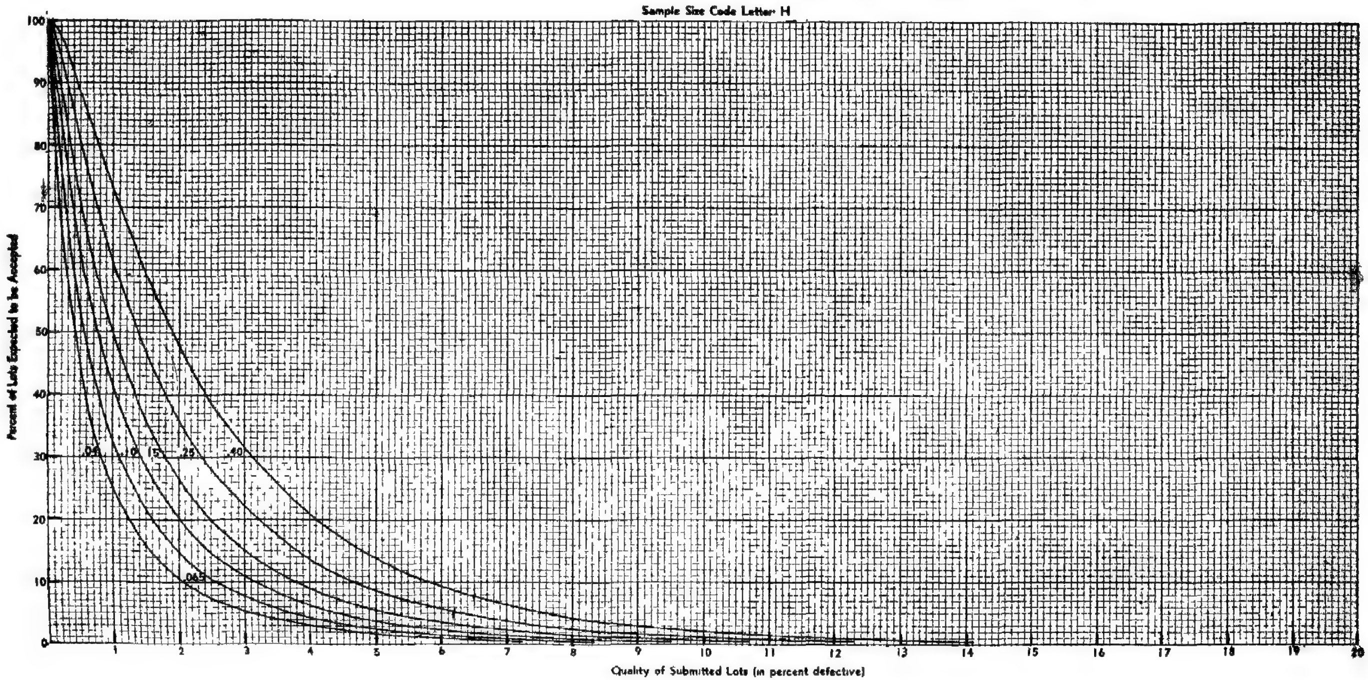


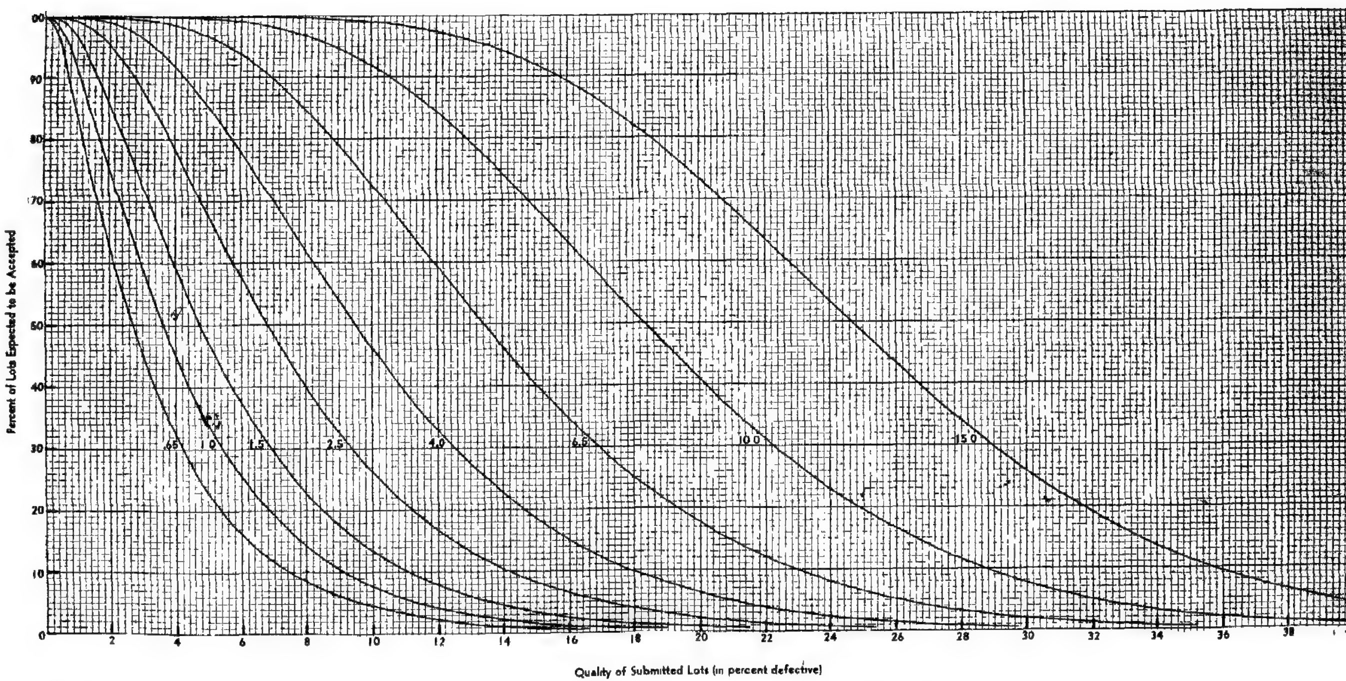
The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.
Note: Figures on curves are Acceptance Quality Levels for normal inspection.

Fig. 7. Sampling Plans for Sample Size Code Letter: H—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)



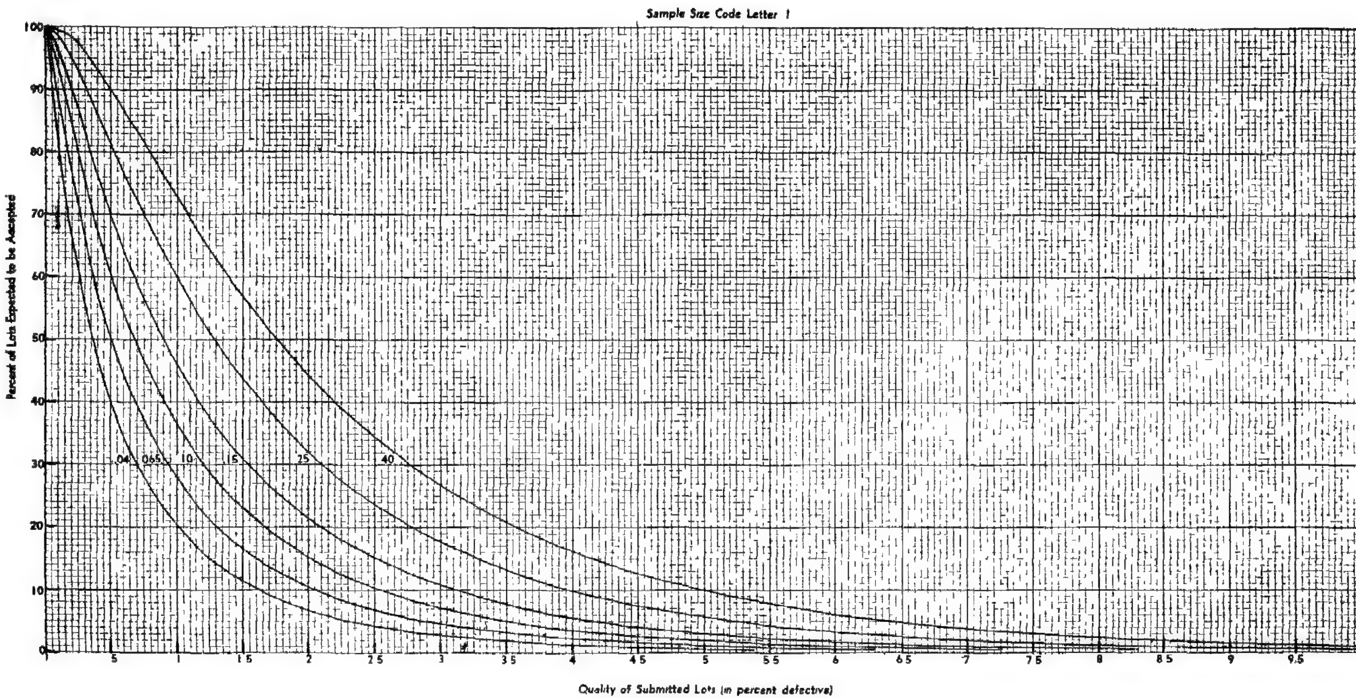


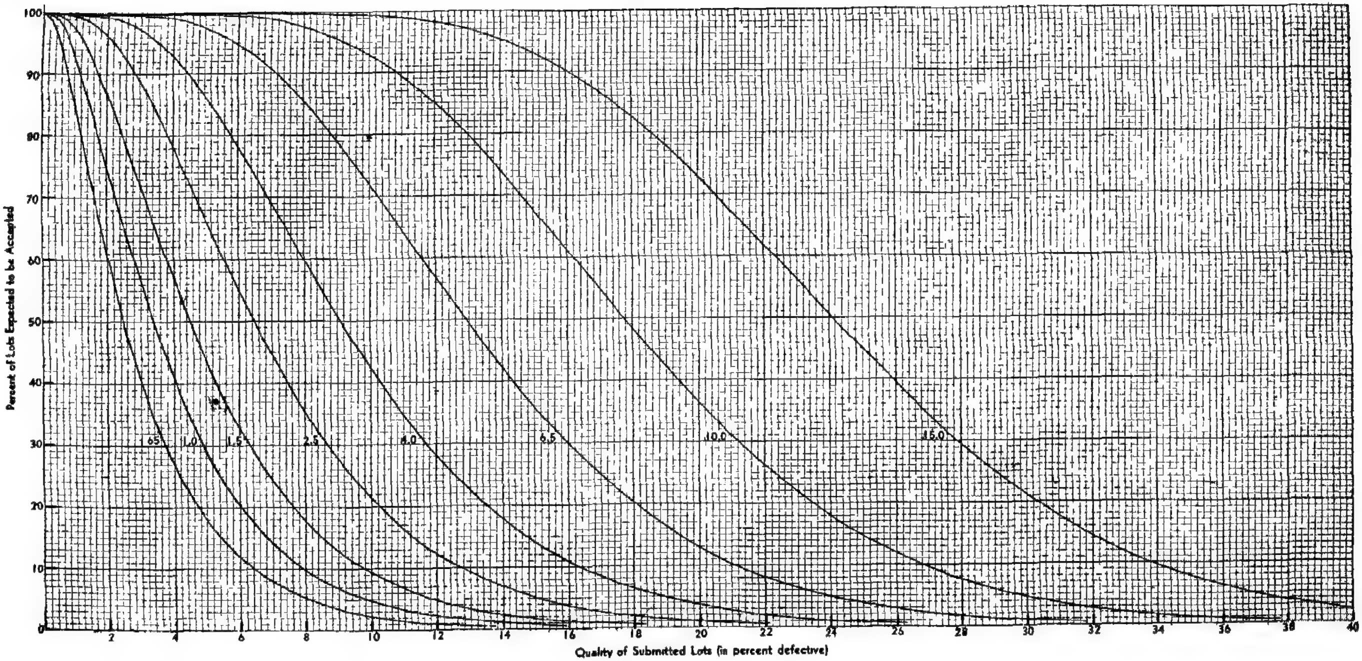
The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.
Note: Figures on curves are Acceptance Quality Levels for normal inspection.

FIG. 8. Sampling Plans for Sample Size Code Letter: I—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)



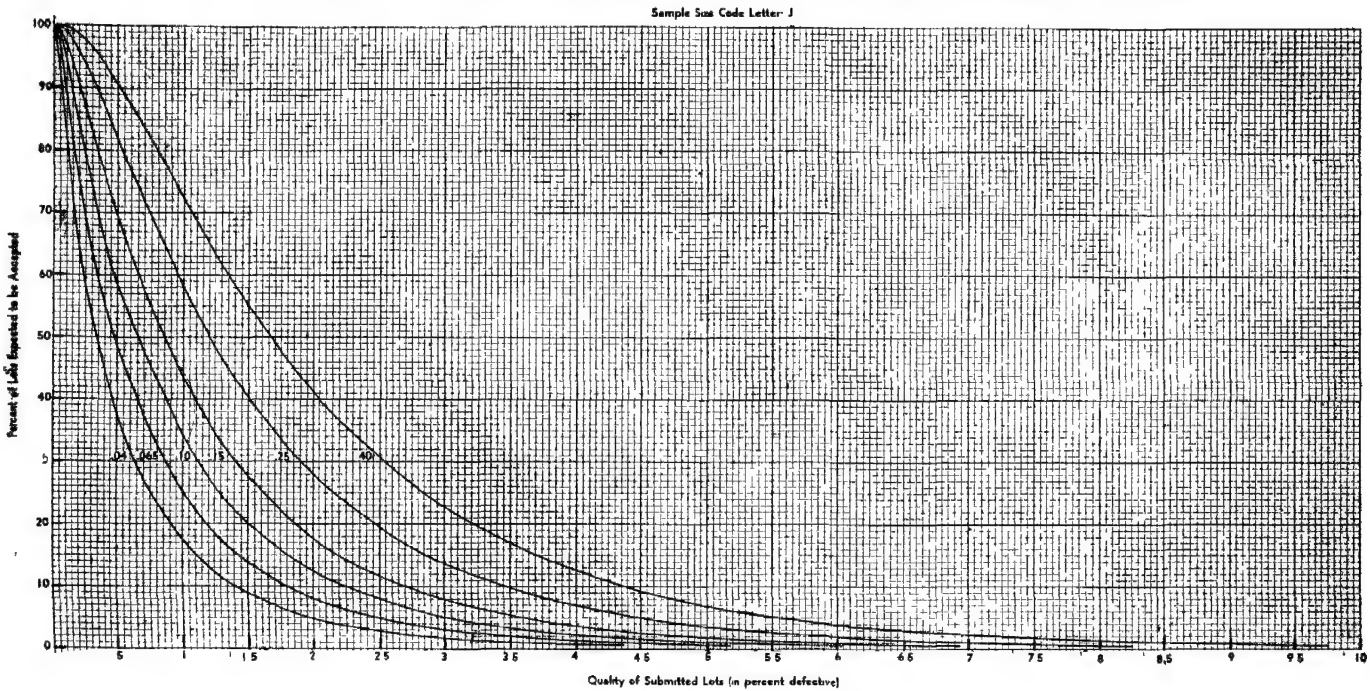


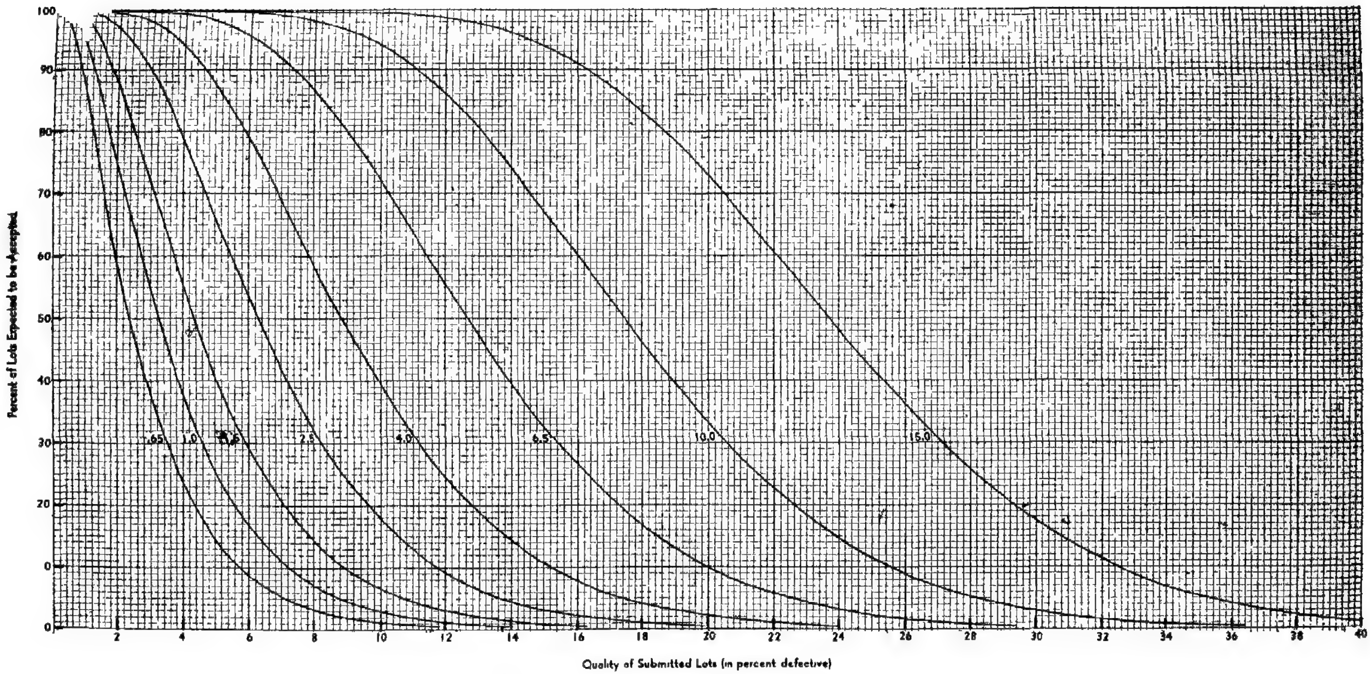
The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.
Note Figures on curves are Acceptance Quality Levels for normal inspection.

FIG. 9. Sampling Plans for Sample Size Code Letter: J—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)



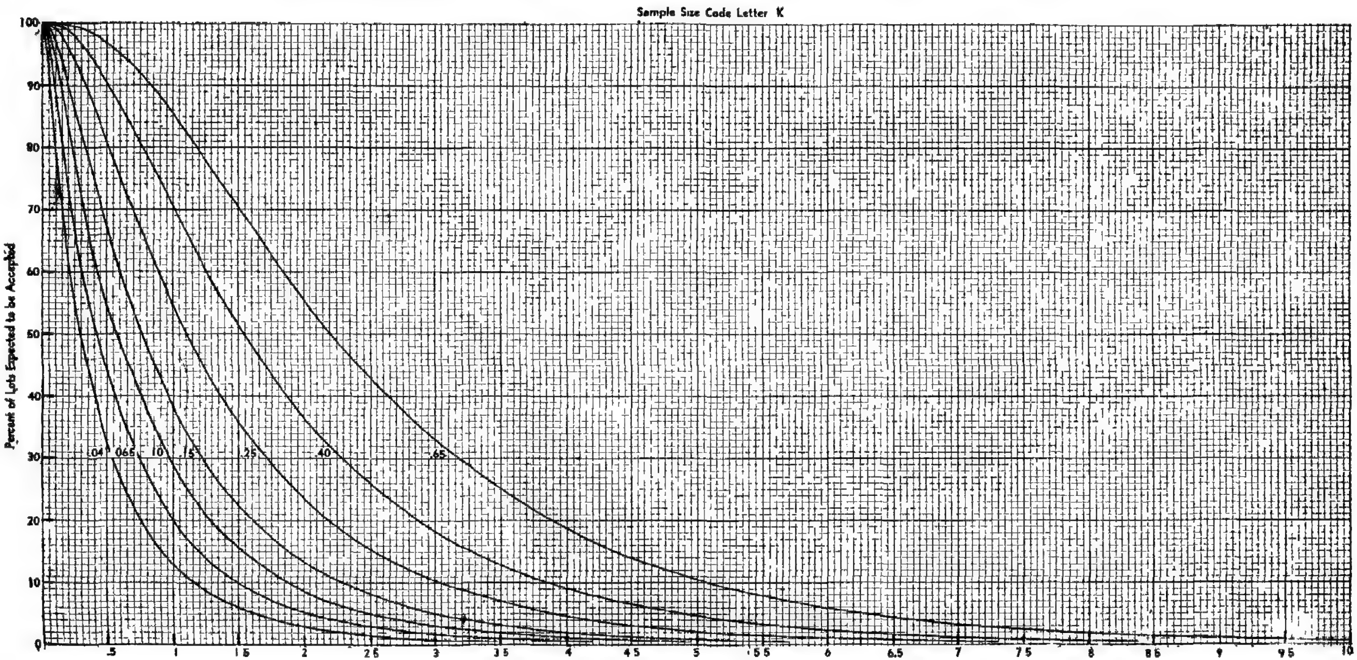


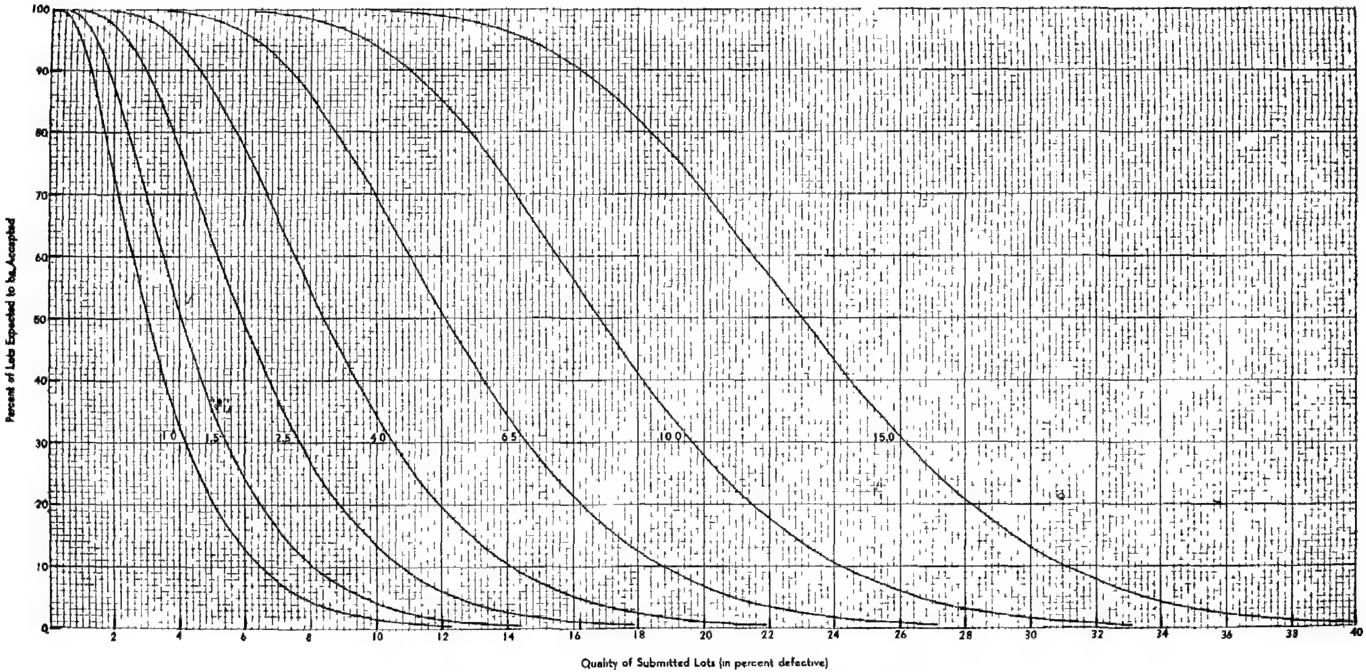
The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.
Note: Figures on curves are Acceptance Quality Levels for normal inspection.

FIG. 10. Sampling Plans for Sample Size Code Letter: K—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)





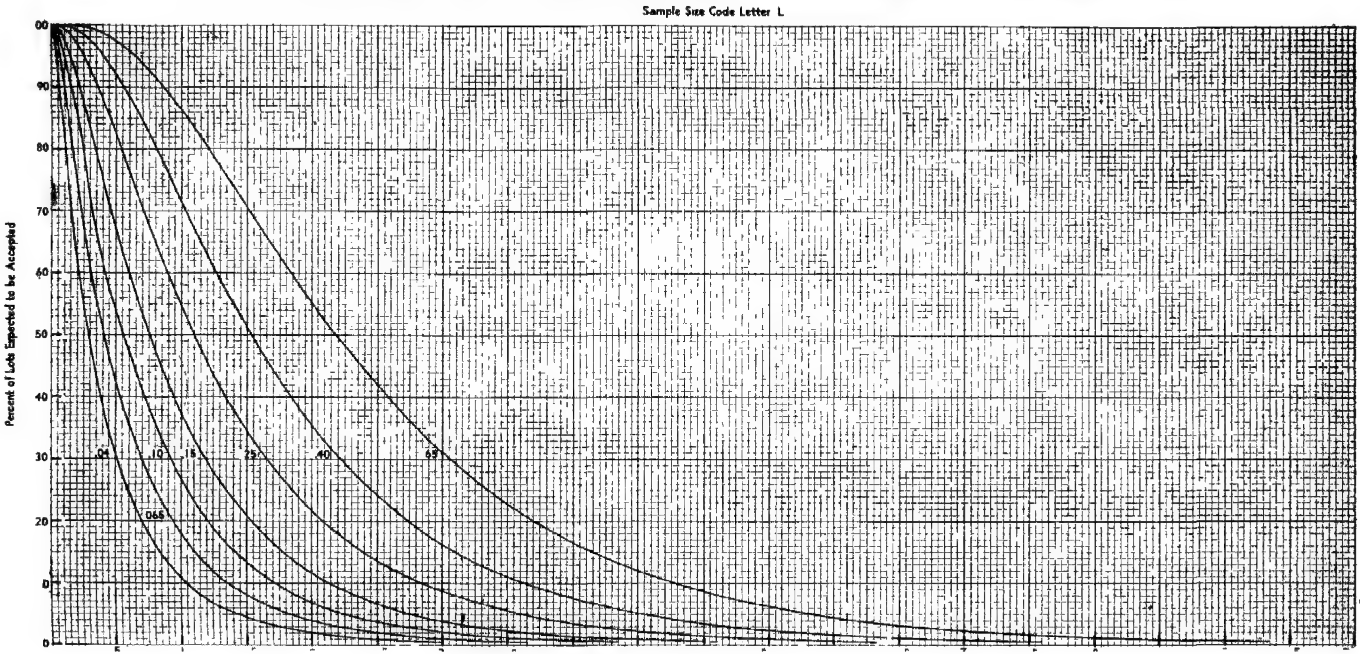
The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution

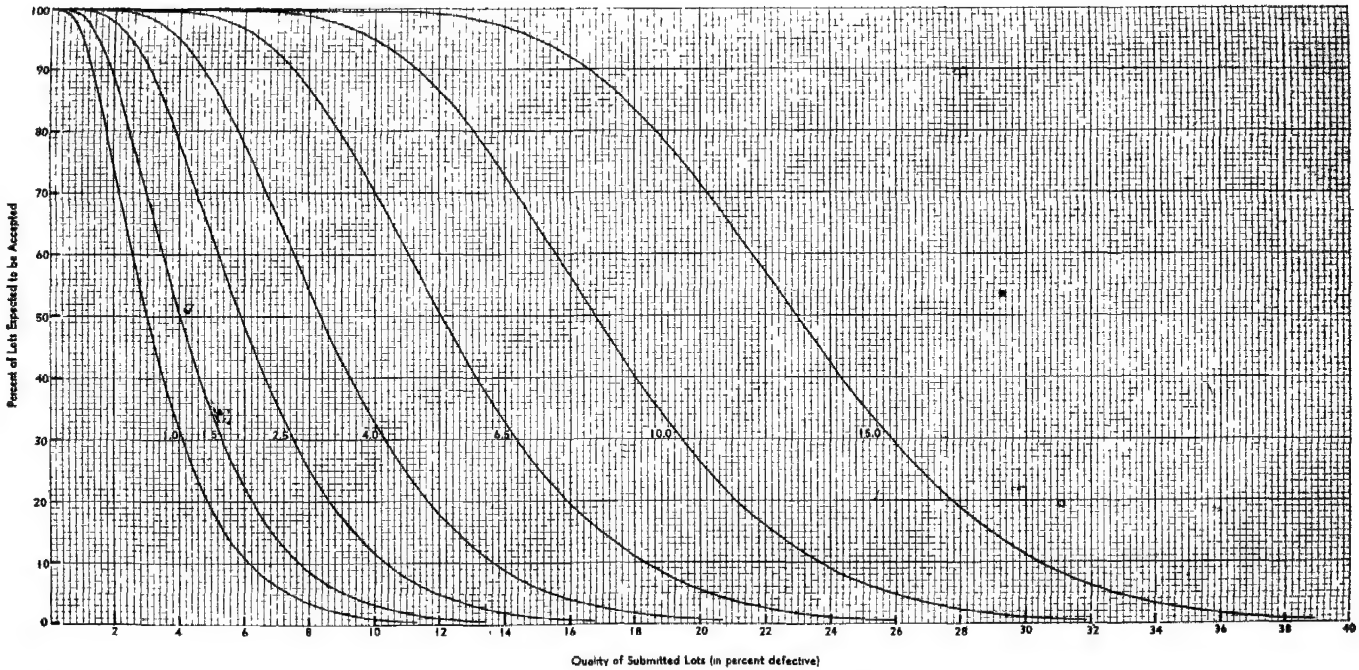
Note Figures on curves are Acceptance Quality Levels for normal inspection

FIG. 11. Sampling Plans for Sample Size Code Letter: L—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)



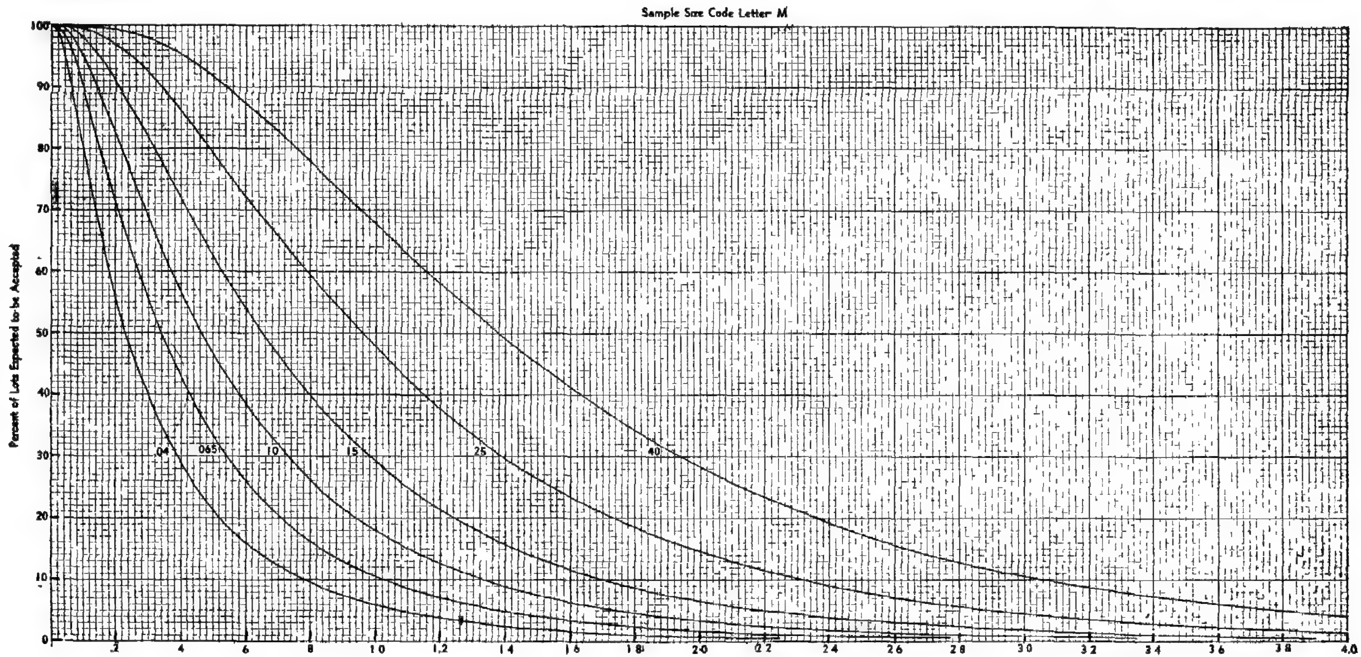


The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.
Note: Figures on curves are Acceptance Quality Levels for normal inspection.

FIG. 12. Sampling Plans for Sample Size Code Letter: M—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)



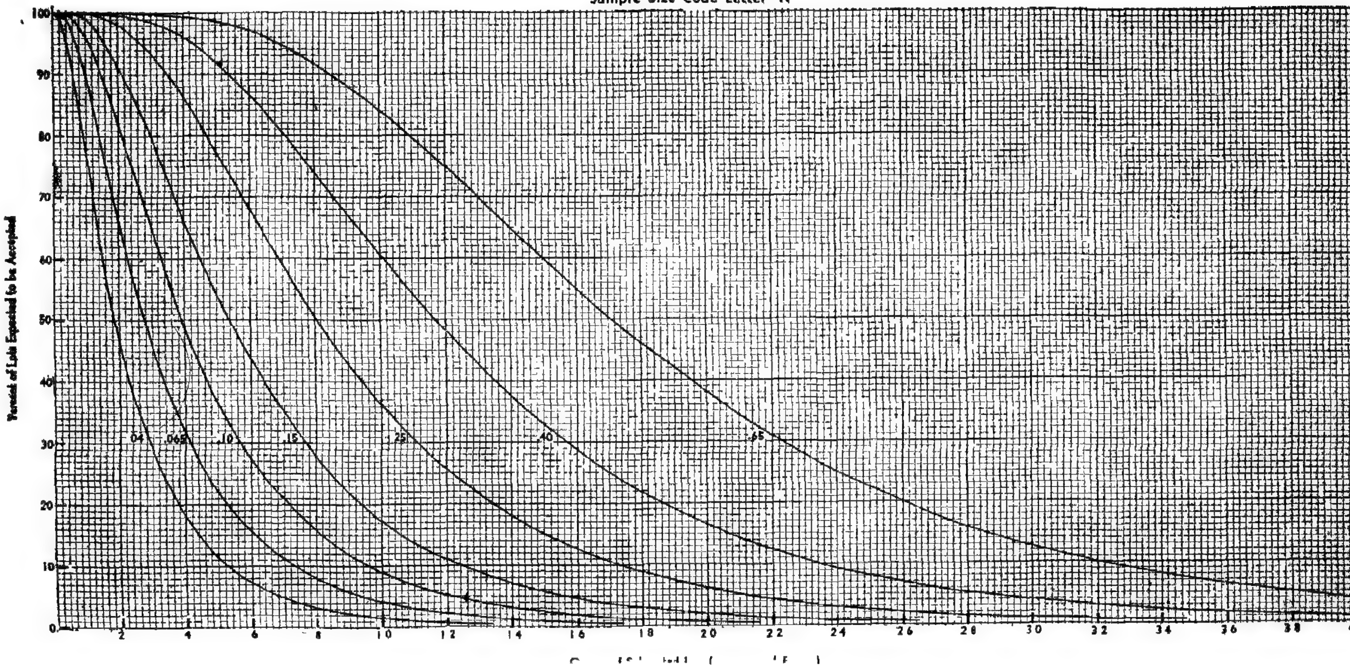
The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution
Note: Figures on curves are Acceptance Quality Levels for normal inspection.

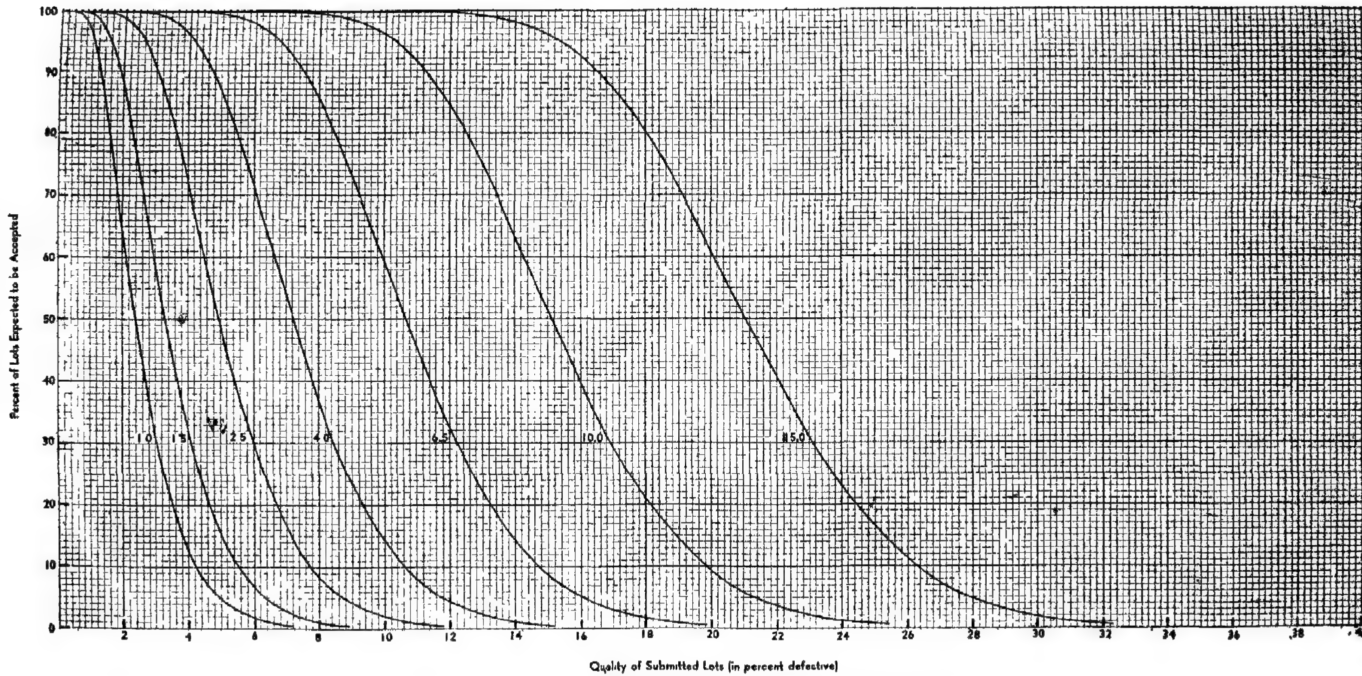
Fig. 13. Sampling Plans for Sample Size Code Letter: N—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)

Sample Size Code Letter: N



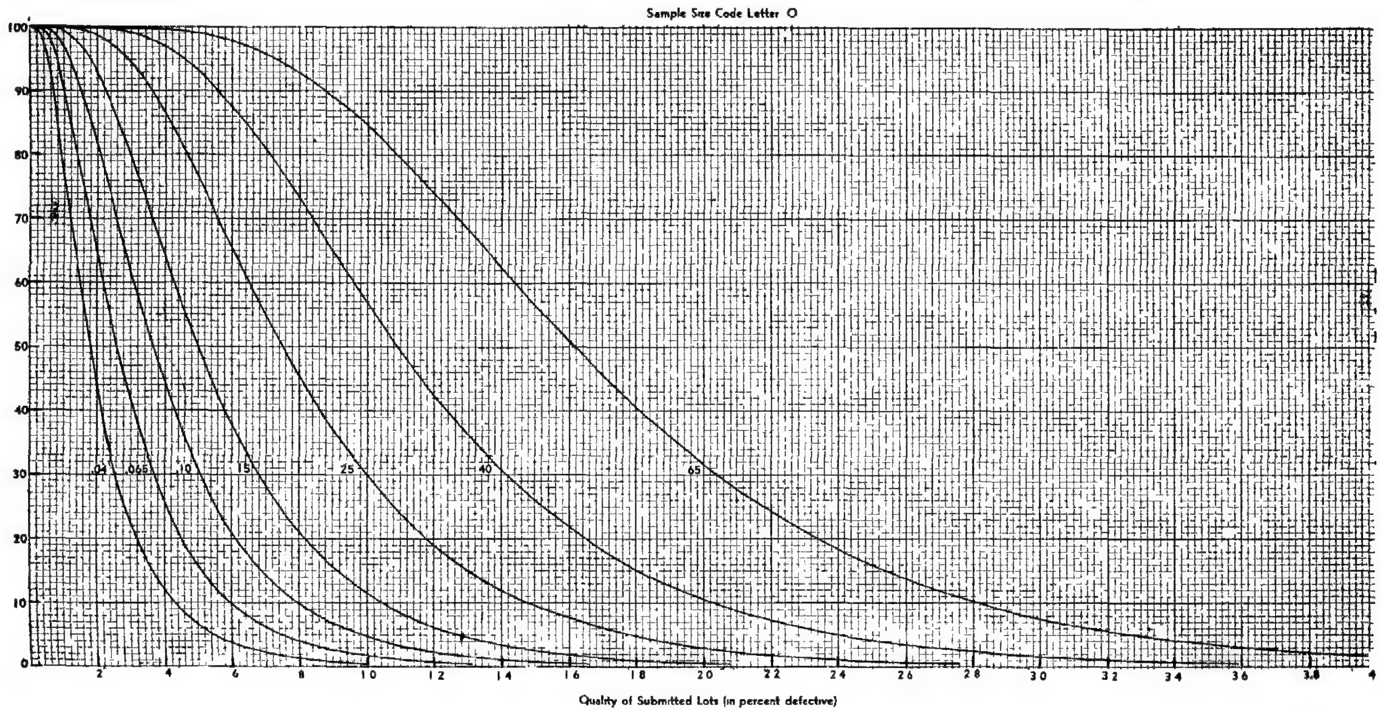


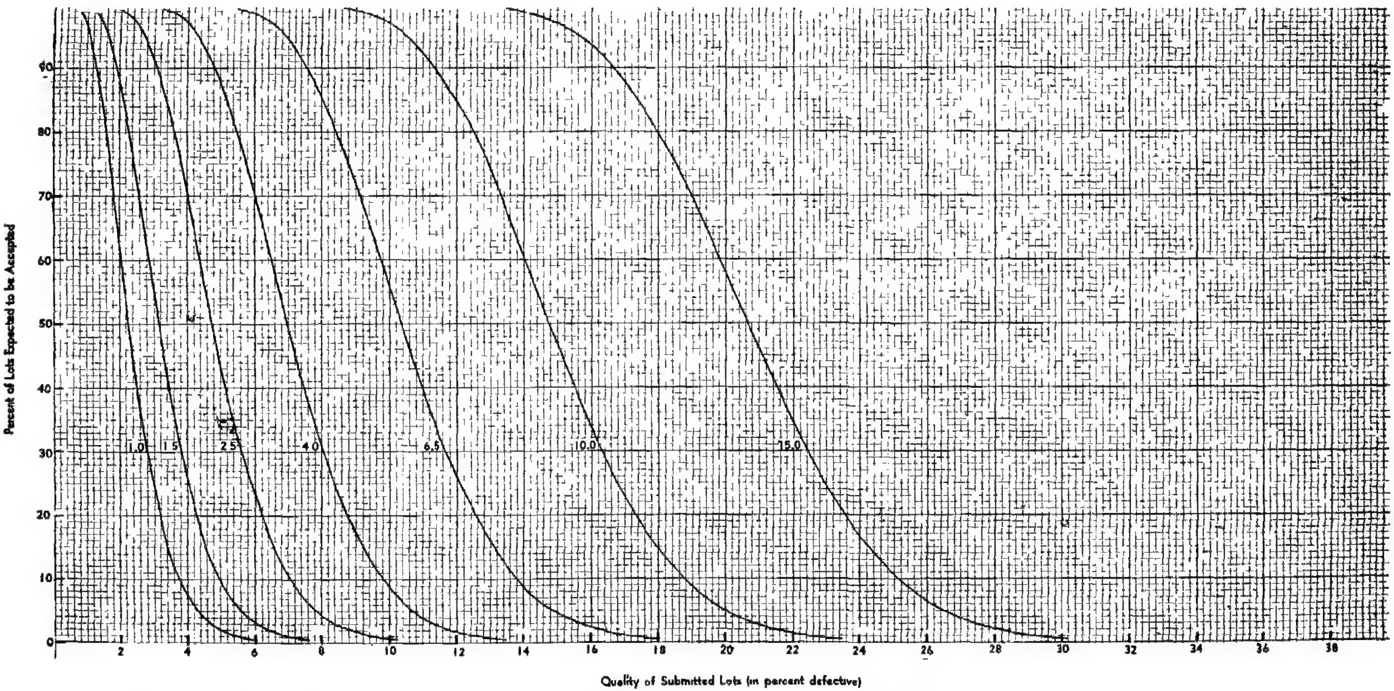
The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.
Note Figures on curves are Acceptance Quality Levels for normal inspection

Fig. 14. Sampling Plans for Sample Size Code Letter: O—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)



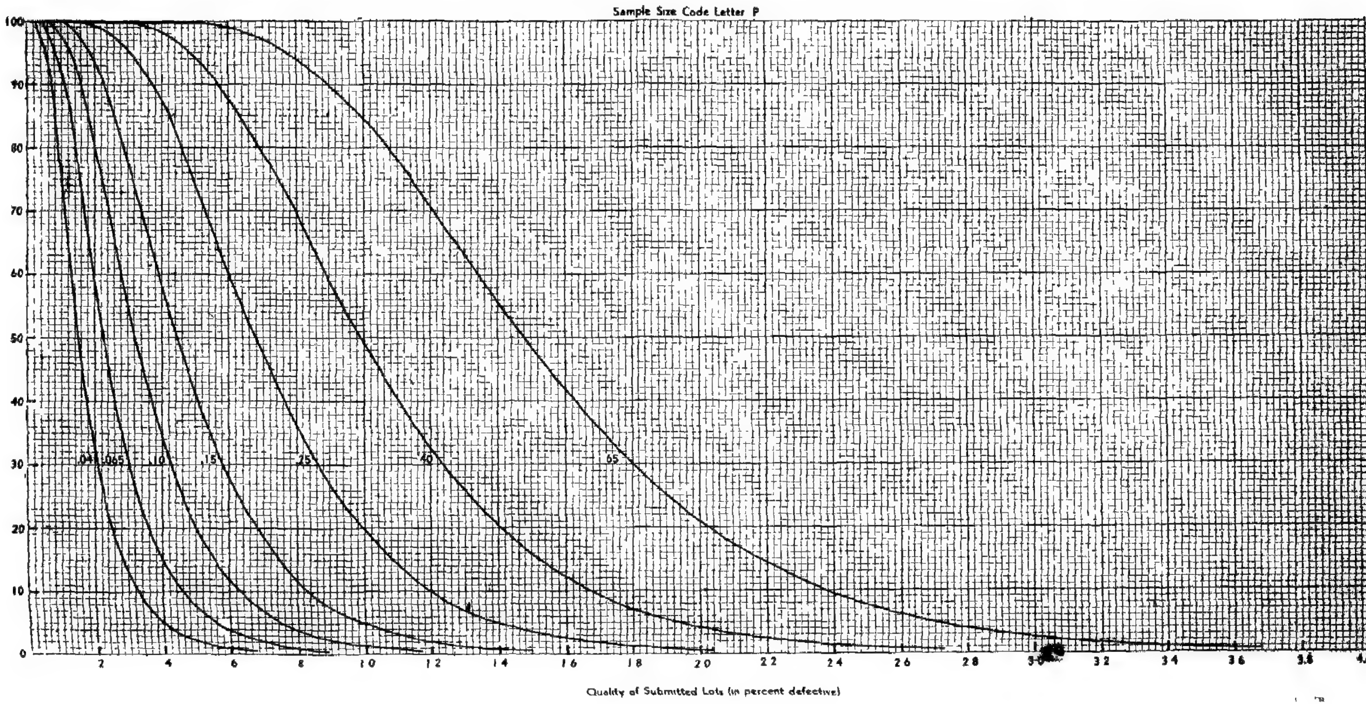


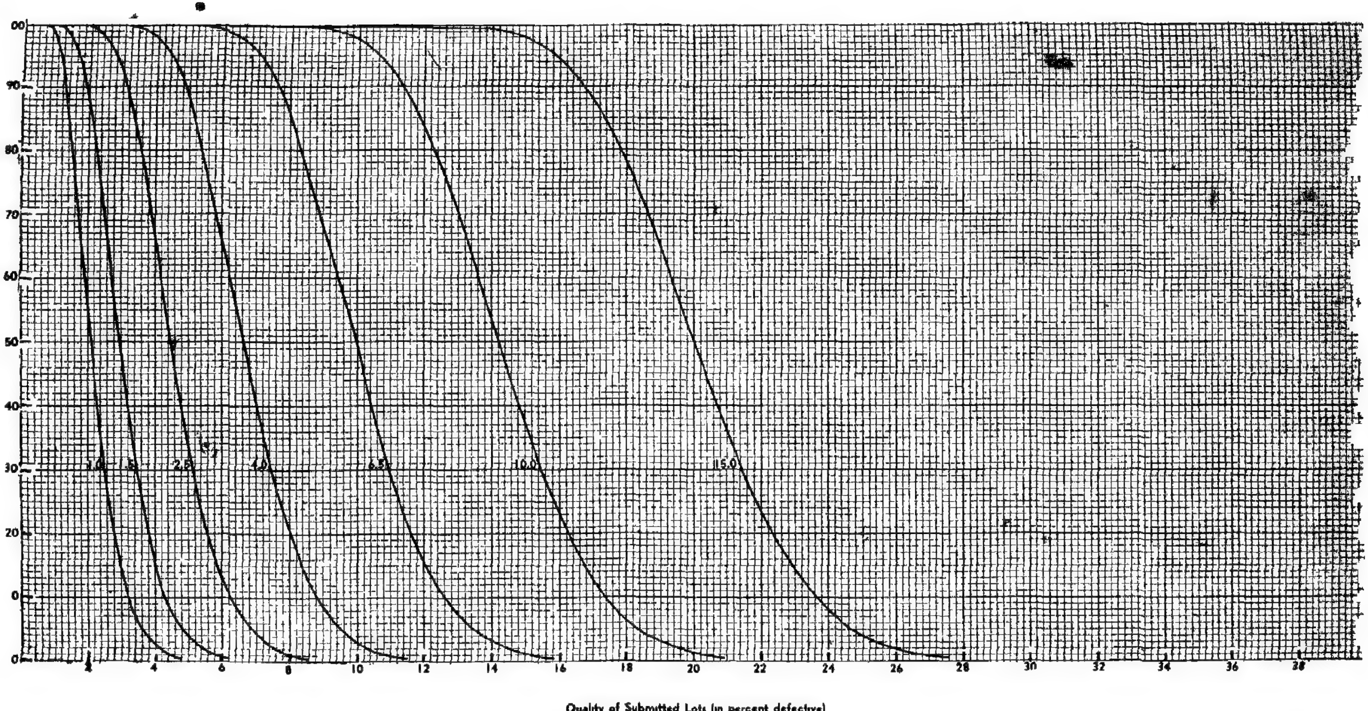
The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.
***Note.* Figures on curves are Acceptance Quality Levels for normal inspection.**

FIG. 15. Sampling Plans for Sample Size Code Letter: P—*Continued*

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviation are essentially equivalent)

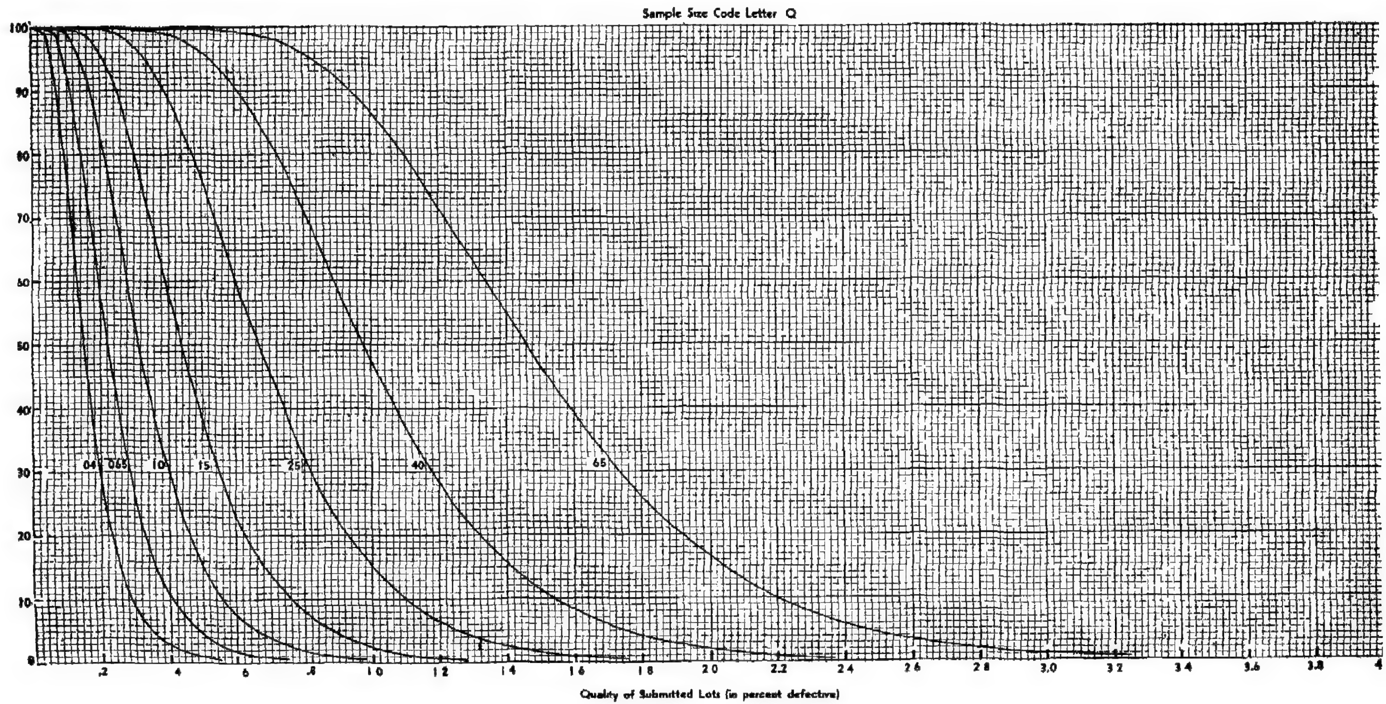


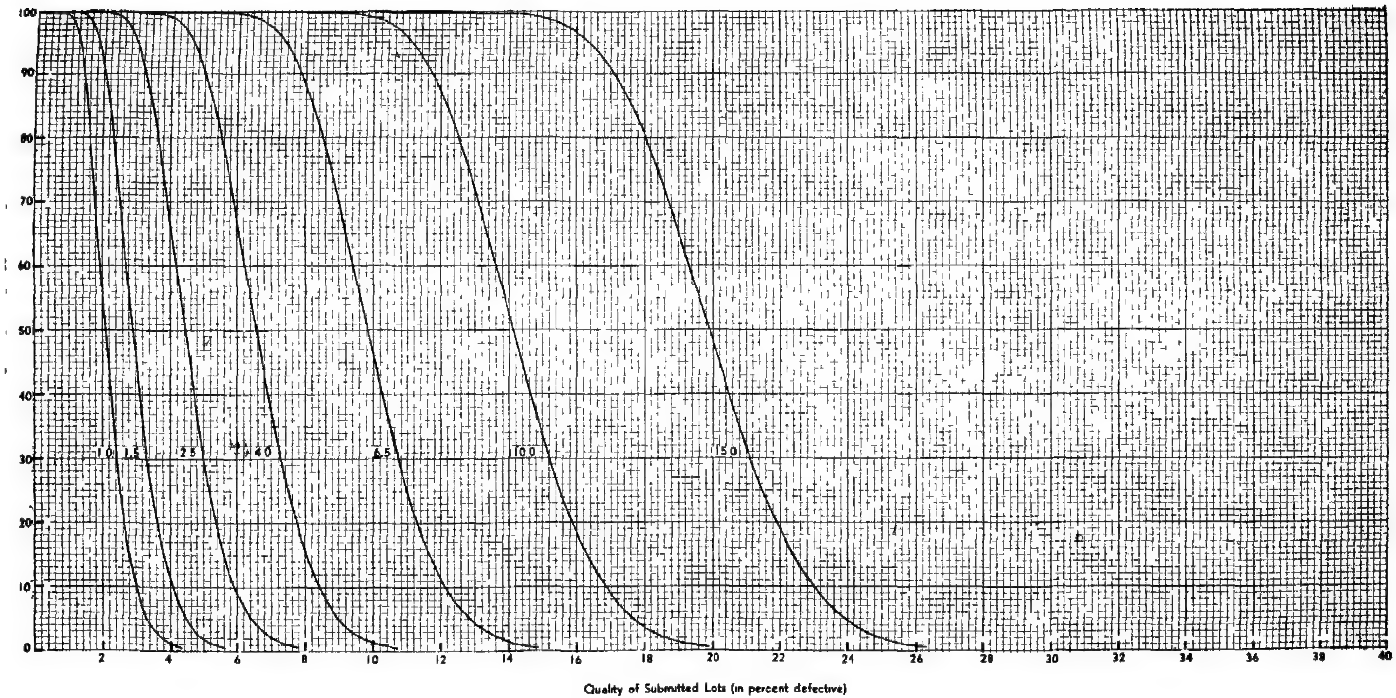


The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.
Note: Figures on curves are Acceptance Quality Levels for normal inspection.

OPERATING CHARACTERISTIC CURVES FOR SAMPLING PLANS BASED ON UNKNOWN STANDARD DEVIATION

(Curves for sampling plans based on average range and known standard deviations are essentially equivalent)





The values of the per cent of lots expected to be accepted are valid only when the submitted items follow a normal distribution.
Note: Figures on curves are Acceptance Quality Levels for normal inspection.

BIBLIOGRAPHY

[1] Blackwell, David, "Conditional expectation and unbiased sequential estimation," *Annals of Mathematical Statistics*, 18 (1947), 105–10.

[2] Bowker, Albert H., and Goode, Henry P., *Sampling Inspection by Variables*. New York: McGraw-Hill Book Company, Inc., 1952.

[3] Craig, Cecil C., "Note on the distribution of non-central t with an application," *Annals of Mathematical Statistics*, 12 (1941), 224–28.

[4] Ireson, W. Grant., *Sampling Tables for Inspection by Variables*, Technical Report No. 7, Office of Naval Research, Contract No N6onr 25126, Applied Mathematics and Statistics Laboratory, Stanford University, (1952). This report may be requested from the Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California.

[5] Ireson, W. Grant, and Resnikoff, George J., *Sampling Tables for Inspection by Variables Based on the Average Range*, Technical Report No. 11, Office of Naval Research, Contract No N6onr 25126, Applied Mathematics and Statistics Laboratory, Stanford University, (1952) This report may be requested from the Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California.

[6] Jennett, W J., and Welch, B. L., "The control of proportion defective as judged by a single quality characteristic varying on a continuous scale," *Supplement to the Journal of the Royal Statistical Society*, 6 (1939), 80–8.

[7] Johnson, N L., and Welch, B L., "Applications of the non-central t-distribution," *Biometrika*, 31 (1940), 362–89

[8] Lehmann, E. L., and Scheffé, Henry, "Completeness, similar regions, and unbiased estimation, Part 1," *Sankhya*, 10 (1950), 305–40.

[9] MIL-STD-105 A, *Sampling Procedures and Tables for Inspection by Attributes*, Washington, D. C: Sup't of Documents, Government Printing Office, 1950.

[10] Patnaik, P B., "The use of mean range as an estimator of variance in statistical Tests," *Biometrika*, 37 (1950), 78–87.

[11] Resnikoff, George J., *A New Two Sided Acceptance Region for Sampling by Variables*, Technical Report No 8, Office of Naval Research, Contract No. N6onr 25126, Applied Mathematics and Statistics Laboratory, Stanford University, (1952) This report may be requested from the Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California

[12] Resnikoff, George J., *The Distribution of the Average Range for Subgroups of Five*, Technical Report No 15, Office of Naval Research, Contract No. N6onr 25126, Applied Mathematics and Statistics Laboratory, Stanford University, (1954). This report may be requested from the Applied Mathematics and Statistics Laboratory, Stanford University, Stanford, California.

[13] Statistical Research Group, Columbia University, *Sampling Inspection, Principles, Procedures, and Tables for Single, Double, and Sequential Sampling in Acceptance Inspection and Quality Control*, New York: McGraw-Hill Book Co., Inc., (1948)

[14] Wallis, W. Allen, "Use of variables in acceptance inspection for per cent defective," Chapter 1 in Statistical Research Group, *Techniques of Statistical Analysis for Scientific and Industrial Research and Production and Management Engineering*. New York: McGraw-Hill Book Co., Inc., (1947), 6–93

[15] Wolfowitz, J., "Confidence limits for the fraction of a normal population which lies between two given limits," *Annals of Mathematical Statistics*, 17, (1946), 483–88.

PRACTICAL APPLICATIONS OF THE THEORY OF EXTREME VALUES*

BRADFORD F. KIMBALL
New York State Public Service Commission

PROFESSOR GUMBEL delivered a series of four lectures on the theory of extreme values at the National Bureau of Standards under the sponsorship of the Applied Mathematics Division in the fall of 1949. An account of these lectures has recently appeared in published form as one of the *Applied Mathematics Series* of the National Bureau of Standards At present this publication represents the only comprehensive source of material on this subject, including a concise account of both the theory and some of its applications.

Professor Gumbel was the first to call the attention of engineers and statisticians in this country to possible applications of the formal "extreme-value" theory to certain distributions which had previously been treated empirically [3, 4] The first type of problem so treated in this country had to do with meteorological phenomena—annual flood flows, precipitation maxima, etc. This was in 1941. Since that time the field of fruitful applications has increased greatly, and it continues to increase.

Someone might ask at this point: "Just what is meant by a *distribution* of extreme values?" Essentially an extreme value is an ordered sample value. Thus a sample of n values x_i is arranged in ascending or descending order of magnitude so that the subscript i indicates order If ordered from low to high x_1 becomes the lowest extreme and x_n the highest extreme. One may also consider the "mth" extreme which refers to the mth ordered value proceeding upwards or downwards from one end of the series If now a *succession* of samples is taken from the same universe, interest may center about a single extreme value of each such sample. In many problems this extreme taken from each sample may be the *only* value recorded—as in maximum annual flood flows, or extinction times of bacteria. The question then arises as to the type of probability distribution which may be expected to apply to such a series of observed extremes It is to the study of the theoretical forms of such

* A review article on *Statistical Theory of Extreme Values and Some Practical Applications*, by E. J. Gumbel, A Series of Lectures, National Bureau of Standards, Applied Mathematics Series, 33 (Washington, D C. U S Government Printing Office, 1954), and *Probability Tables for the Analysis of Extreme-Value Data*, National Bureau of Standards, Applied Mathematics Series, 22 (Washington, D. C.. U S. Government Printing Office, 1953)

distributions and their applications that Gumbel's lectures are addressed.

The published lectures cover 51 pages including a bibliography of some 80 references. The material is arranged as follows: *Lecture 1*, Survey of Practical Applications of Extreme-Value Theory, *Lecture 2*, Exceedances, Return Periods and Probability Papers; *Lecture 3*, Exact and Asymptotic Distributions of Extremes; *Lecture 4*, Applications. Thus, the theory of extreme-value distributions is not treated until the third lecture. The reason for this is that certain of the methods developed to handle extreme-value data are general and can be applied to other types of distributions These methods are accordingly introduced first, with the thought that they can be applied to more general types of distributions.

It will possibly clarify the comments which are to follow if at this point the specific form of the extreme-value distribution which furnishes the *pièce de résistance* of these lectures, is set forth This is the extreme-value distribution of Type I (see (3 16) and (3 17) of the Lectures). It is the asymptotic form of the exact distribution of maximum values in sample from a universe whose distribution is such that its upper tail declines to zero like an exponential Normal and chi-square distributed universes are of this type. It is sometimes referred to as the doubly exponential distribution of maximum values since it is the only one of the three types which is a double exponential.

The probability density function ϕ and the cumulative probability Φ are usually expressed by means of the equations

$$(1) \qquad \Phi(y) = e^{-e^{-y}}, \qquad \phi(y) = e^{-y-e^{-y}} \qquad \infty < y < \infty$$

$$(2) \qquad y = \alpha(x - u)$$

where the parameters are α and u [5] The intermediate variable y, is often referred to as the "reduced" variate and α or $1/\alpha$ as the "scale" parameter since it relates the scale of measure applicable to the observed variate x to that of the intrinsic variable y. The parameter u turns out to be the *mode* of the extreme-value distribution. Hereafter when we refer to "extreme-value" distribution without indication of type we shall mean the distribution (1) above.

For purposes of graphical fitting, extreme-value plotting papers have been prepared so that with x measured along one axis *distances* along the other axis are proportional to y, but *readings* along the y-axis are the values of Φ given by (1) Obviously then the critical relation (2) which determines the parameters, is a straight-line relation on this plotting paper. Most of the charts illustrating applications in Lecture 4

are based on (1) and (2) above. They deal with a series of extreme values which itself is completely ordered as to size The curious problem then arises as to how the plotting points x_m, $\Phi(y_m)$ should be chosen in order to represent the data fairly. The simplest convention is to employ the general relation applicable to any distribution

$$(3) \qquad E[\Phi(y_m)] = \frac{m}{n+1}$$

and accordingly take Φ equal to $m/(n+1)$ for an observation x_m. This convention might be referred to as the convention of "equidistant spacing" since $\Phi(y_{m+1}) - \Phi(y_m) = 1/(n+1)$ for any value of m. It is the method preferred by Gumbel [2, pp 13–15]

Another choice of the plotting positions of $\Phi(y_m)$ is to use the relation

$$\hat{\Phi} = \Phi[E(y_m)]$$

which has the advantage that it eliminates a bias when the line is fitted by eye This method is more complicated in that $E(y_m)$ has to be taken from a table or chart Such a table and chart have been prepared, but are not yet available in the published literature [10].

One may consider the graphical fitting as a preliminary step to be followed by an analytical method of determining the parameters. Viewed in this light the simpler choice, using the relation (3), seems adequate Where extrapolation and forecasts are important, analytical procedures of fitting the parameters are required

Two such methods of determining the parameters are suggested by Gumbel in his lectures In [3] he suggests that the method of moments be used This results in the relations

$$(4) \qquad \frac{1}{\alpha} = \frac{\sqrt{6}}{\pi} s_x, \qquad u = \bar{x} - \gamma/\alpha$$

where $\bar{x}$ and s_x denote the mean and standard deviation of the observed values of x, and γ denotes Euler's constant [2, (3 29) and section on *Estimation of Parameters*]. These are derived directly from the mean and standard deviation of the reduced variable y, which are

$$(5) \qquad \bar{y} = \gamma, \qquad \sigma = \pi/\sqrt{6}$$

Later Gumbel derived intuitively a modified form of this procedure,

given by

$$\frac{1}{\alpha} = \frac{s_x}{\sigma(n)}, \qquad u = \bar{x} - \frac{\overline{y(n)}}{\alpha} \tag{6}$$

where $\overline{y(n)}$ denotes the mean value of y found from the n plotting positions $\Phi(y_m) = m/(n+1)$, and $\sigma(n)$ denotes the standard deviation of these values of y [2, Lectures 2 and 3] Gumbel points out that as n becomes infinite $y(n)$ and $\sigma(n)$ approach the mean and standard deviation for the continuous case, if such exist He also points out that the above relation (6) is general in character, applying to any distribution such that the cumulative probability is a known function of an intermediate variable y, which in turn is an unknown linear function of the observed variable x, as in (2) above

This is the method of determining the parameters which is recommended by Gumbel, as indicated in the *Summary of Procedures* on page 46. The quantities $\overline{y(n)}$ and $\sigma(n)$ depend upon the sample size n. On page 29 of the lectures may be found a table of these quantities for various sample sizes, and directions for interpolating between tabulated values A table applicable to fitting the normal distribution by this method is also shown on page 17.

The derivation of (6) is set forth in the section *Fitting Straight Lines* in Lecture 2 Having plotted the data using convention (3), Gumbel sets up the moment conditions of least squares for minimizing vertical distances and horizontal distances, thus arriving at two sets of estimates of the parameters $1/\alpha$ and μ (the parameter μ is used in Lecture 2 where the distribution function is not necessarily that of extreme values). He then takes the geometric mean of these two sets of estimates and in this way arrives at the relation (6) above, which appears as (2 20) and (3 39) in the lectures

One might note that the relation (6) is the relation which would be obtained if the line were fitted to the *normalized* variables $(y-\overline{y(n)})/\sigma(n)$ and $(x-\bar{x})/s_x$ by minimizing the sum of the squares of the *perpendicular* distances from the plotted points to the fitted line

The method does not consider fundamental questions of bias and efficiency. Although these lectures were delivered to an audience interested primarily in applications, one should call attention to the fact that the above method is a *tour de force* in the interests of producing a simple, general method which will appeal to the practical man, and be sufficiently accurate for his purposes. Since the method is recommended in the *Summary* for use in fitting the distribution of extreme values we refer the reader to an investigation made by Lieblein [11,

p. 48] in which he concludes from empirical sampling tests that what he calls "the original Gumbel estimator" (referring to 3.39) of the Lectures and (6) above) "is more biased and much less efficient than the simplified form" referring to (3.29) of the Lectures and (4) above).

One might remark that this whole problem of fitting the extreme-value distribution has proved to be a knotty one. The efficient, maximum-likelihood estimate of the parameters is somewhat complicated, especially for large samples [8]. Another method [8, (4 5)] retains the qualification of sufficiency in the extended sense, but is more or less complicated In the interests of simplicity Lieblein has introduced a method applicable to samples of any size based on six (or less) order statistics which he finds more efficient than the Gumbel method, using (4) above This has an efficiency ratio of about 80 per cent relative to the maximum likelihood estimate for theoretical x (defined more precisely as the Cramèr-Rao lower-bound), where the cumulative probability is 0.99 In terms of annual flood estimates this means that the estimate of annual flood which is expected to be reached or exceeded 1 per cent of the time has an efficiency ratio of 80 per cent [11, Table IV]

One may conclude that for many practical purposes the method of plotting indicated by (3) above which is recommended by Gumbel, and the method of fitting indicated by Gumbel's equations (3 29) and by (4) above, will be satisfactory At the time the lectures were given Lieblein's work on the use of selected order statistics as estimators was not available. This should be looked into if forecasting is critical. If an intensive job on a single series is to be done, it would be well to compare results obtained by the method of maximum likelihood or the similar method which also uses "sufficient statistical estimation functions" [8, (4 5) and (5 4)]

The principal warning signal which this reviewer believes should be set up for the lay reader is in connection with Gumbel's so-called "control curves" These control curves are based for the most part on the formulas

$$\begin{aligned} \sqrt{n}\,\sigma(y_m) &= \sqrt{\Phi(y_m)\left[1 - \Phi(y_m)\right]}/\phi(y_m) \\ \sigma(x_m) &= \sqrt{n}\,\sigma(y_m)/\sqrt{n}\,\alpha \end{aligned} \tag{7}$$

where y_m denotes the reduced variate of the mth ordered value, x_m the mth ordered value observed, and α the population parameter (see sections entitled *Control Curves* on pp 17, 27, 31 and 48) This means that Gumbel prefers to ignore the sampling error involved in

estimating the *position* of the fitted distribution curve (straight line on probability plotting paper).

For the distribution of maximum values [11, pp 27 and 59] Lieblein has derived the sampling variance of the position of the curve for an estimate based on order statistics The asymptotic variance (applicable to large samples) for the efficient (maximum likelihood) method of estimate has been found by Kimball *[9, (12)]*.

Apparently Gumbel does not wish to include the sampling variance involved in the estimation of the parameters and the resulting sampling variance in the position of the fitted distribution curve This means that the experimenter should use the control curves only as a check on the amount of *scatter* of the data points about the fitted curve Indirectly, a bias of scatter on one side or the other might indicate error in the estimated position of the distribution curve. It might be noted here that the theoretical measures of the scatter about the theoretical distribution curve—see (2 25) and (2 26)—which define Gumbel's control curves, are based on intrinsic properties of the distribution, and on the fitted data only indirectly through the involvement of the parameter α Thus the control curves are a measure of the inherent dispersion to be expected of individual observations about their theoretical means An analytical test of goodness of fit would be based essentially on a comparison of actual behavior of deviations relative to such expected behavior (for example, the test devised by Sherman [12]).

The warning signal goes up when the experimenter thinks that the control curves offer a confidence band for purposes of extrapolation or estimation of what might be expected to happen at some specific value of the independent variable (usually time) Researchers working with the conventional least squares fitting of polynomials know that to the variance of scatter is *added the sampling variance of the position of the fitted curve* at the point in question. An allowance of the same sort will have to be made in dealing with the extreme-value distribution, the allowance depending upon the extent to which the estimators of dispersion about the fitted line are statistically independent of estimators of the position of the theoretical line Until this is done the complete treatment of the variance of a forecast remains an unsolved problem

The lay reader might well start with Lecture 4 and work backwards by first referring to the section *Summary of Procedures* and then to explanations of the methods which appear in the earlier lectures.

In detail, the subject-matter covered by the lectures is as follows. Lecture 1 is a sort of general introduction designed for the lay reader

and should tend to stimulate his curiosity. This reviewer finds the accounts of the history of the theory, and the historical aspects of the attacks on such problems as flood frequencies and breaking strength of materials, of great interest. The author occasionally lets his enthusiasm run away with him, however, as in the statement " . . the distributions of floods, long studied by engineers, can be fully understood through this theory".

Lecture 2 opens with an introduction to the distribution-free law of exceedances This is followed by a brief exposition of some of Gumbel's work on "the law of rare exceedances" with application to tolerance limits A more complete treatment is to be found in the article by Gumbel and von Schelling [6]

The next four sections *The Return Period, Expected Extremes, Construction of Probability Papers* and *Plotting Positions* involve valuable expository material. These four sections, and indeed the rest of Lecture 2, are designed to apply to an ordered sample of data taken from any distribution where the cumulative distribution function $F(x)$ is a completely defined monotonically increasing function of x over its range from $F=0$ to $F=1$ The concept of "the return period" defined as $1/(1-F(x))$ and denoted by $T(x)$ is introduced. This concept Gumbel apparently believes may be useful in studying many distributions other than that of extreme values His statement following (2 8) which reads: "It is the number of observations such that, on the average, there is one observation equalling or exceeding x" might be better phrased as "*it represents the sample size* such that on the average there would be *just* one observation equalling or exceeding x "

Closely allied to the concept of the "return period" is that of an "expected extreme " This is defined as the value $x=u_n$ for a sample of size n such that the return period is precisely equal to the size of the sample Thus

$$(8) \qquad T(u_n) = n = 1/(1 - F(u_n)), \qquad F(u_n) = 1 - 1/n.$$

This concept has been found useful in the analysis of the transition from an exact distribution of extreme values to the asymptotic distribution (see pp. 19 and 21).

The sections on *Construction of Probability Papers* and *Plotting Positions* point up the problem which arises in determining the proper plotting of the frequencies of completely ordered sample data, for the purposes of graphical methods of curve fitting. The five "postulates" concerning the appropriateness of a plotting scheme, which are listed on page 14, should be taken with a grain of salt. For example postulate

3—"The observations should be equally spaced on the frequency scale"—would not follow under the scheme of using $F(\bar{x}_m)$ as the proper plotting value of the cumulative frequency F for the mth ordered observation x_m where $\bar{x}_m$ denotes $E(x_m)$ In this case $F(\bar{x}_{m+1})-F(\bar{x}_m)$ is *not* in general independent of m, and there seems to this reviewer no inherent reason why this should be required

The section on *Fitting Straight Lines* takes up the problem of the analytical determination of the two parameters involved, and has already been discussed. It should be noted that as introduced in Lecture 2 the method proposed is to apply to *any* distribution $\Phi(y)$ where Φ is known, but y is a linear function of x involving the two parameters to be determined.

The section entitled *An Unsolved Problem* is devoted to the discussion of a linear relation which Gumbel has discovered empirically, which makes it possible to use tables and charts for short-cutting the computation of the quantities $\overline{y(n)}$ and $\sigma(n)$ which appear in the estimation equations (2 20), as applied to the fitting of a normal distribution. This relation is discussed again under the same section heading on page 29 where application is made to the distribution of extreme values.

Lecture 2 concludes with a section *Control Curves* As mentioned previously in this review, these are confidence bands indicating the degree of scatter of the plots of observed sample data which is to be expected about the theoretically correct mean of each such ordered sample value Gumbel points out that near the extremes the distribution should approximate that of the extreme-value distribution more closely than that of the normal distribution and accordingly suggests that another formula be used in measuring this scatter in the neighborhood of the extremes of the data series This is gone into by Gumbel in the later section *Extension of Control Curves* on page 27 and again in the section *Control Curves* on page 31 Lieblein has criticized the Gumbel choice of the parameter α_n, as equal to α in this extension process (see p 65 of [11]). Lieblein, however, is thinking of the control curves as designed to include the sampling variance of the *position* of the fitted curve In terms of the concept that the control curves indicate only the expected scatter about the true position of the fitted curve, Gumbel's conclusions stated in (3.44) and (3 44a) are correct.

Lecture 3 contains the meat of the extreme-value theory The mathematical analysis leading up to the three forms of the extreme value distribution (labelled I, II, and III, pp 21–22) is not easy to follow under

anybody's exposition. This is contracted into three pages, and hence the intuitive approach followed by Gumbel can be forgiven. Indeed if the mathematically inclined statistician should supplement this by studying the Fisher-Tippett and von Mises' articles, Gumbel's approach should aid the conceptual appreciation of the transition from the exact distribution to the asymptotic distribution. One might mention a slip in the explanation of the Fisher functional equation on page 21

$$F^n(x) = F(a_n x + b)$$

Keeping to the notation used in the equation, Fisher assumes n samples of size N (rather than N samples of size n) in order to compare the distribution of the nth extreme among the extremes of the n smaller samples with the extreme of the consolidated sample Thus $F(x)$ here refers to the cumulative distribution of maximum values, usually denoted by $\Phi(x)$ in these lectures

The section *Estimate of Parameters* must be considered at present out of date in coverage The relations (3 29) constitute Gumbel's early preference. As mentioned above, Lieblein finds these preferable to the method of (3 39) recommended in the Summary.

The section *Extreme mth Values* introduces the asymptotic distribution of the mth largest value A curious fact about the distribution of the *largest* extreme value is that the asymptotic variance about its mean is equal to the variance of the parent extreme-value distribution

The section *Return Periods* is of decided interest, introducing an ingenious asymptotic formula (3 48)

$$P(K) = e^{-1/K} - e^{-K}.$$

In terms of a return period T corresponding to an extremum x, this relation gives the probability that an extremum as large as x should occur for the first time after t_1 trials and before t_2 trials are made The formula shows that $t_1 = T/K$ and $t_2 = KT$ where the constant K is determined directly from the probability and is *independent* of T. This formula is recommended in the Summary in connection with the extrapolation of the control curves. Again the reader should be cautioned that the value of T in its relation to x involves a sampling error and hence the interval measures only the degree of scatter to be expected about the *true* $T(x)$.

The present reviewer finds the last section, *Weibull's Use of the Third Type*, of particular interest for several reasons This is of the limited type which he at one time studied as a possibility in describing the

distribution of annual floods. Shigeo Kase [7] has found that the distribution of tensile strengths of rubber can be described by the distribution (3 23) with

$$y = (S - S^*)/\beta$$

where S represents the measure of the tensile strength of specimens with a constant cross section and S^* the median. If one places the lower limit of S at zero one might replace S by $k \log x$ which reduces the distribution precisely to (3 25) with $\omega = 0$. Cumulating the frequency in the opposite sense,

$$\Phi = \exp -(x/v)^k$$
$$\log \Phi = -(x/v)^k$$

This form of the distribution may be studied by the use of log log plotting paper along with the following simple formula for $E[\log \Phi_m]$

$$E[-\log \Phi(y_m)] = \frac{1}{m} + \frac{1}{m+1} + \cdots + \frac{1}{n}$$

which applies to any well-mannered cumulative distribution function. Furthermore, the analytical process of determining the parameters v and k would be considerably simplified [1]

The rest of Lecture 3, not already referred to in the review of Lecture 2, is expository material comparing characteristics of the doubly exponential extreme-value distribution with the normal distribution.

Lecture 4 is an excellent account of the application of the extreme-value theory to many practical problems of a diverse nature such as floods, maximum gust loads in aeronautics, old age, bacteria extinction times, radioactive emissions and breaking strength of materials It is well illustrated and fairly easy reading.

This series of lectures gives a very condensed account of a great deal of pioneer work. Professor Gumbel is to be congratulated on the job that he has done.

Attention should be called here to the recent publication of *Probability Tables for the Analysis of Extreme-Value Data* [13]. These tables should certainly be acquired by anyone doing work involving the extreme-value distribution

This publication contains the following six tables:

Table 1. Cumulative probability and density functions of extremes.

Table 2. Inverse of the cumulative probability function of extremes.

Table 3. Probability density of extremes as a function of the cumulative probability.

Table 4. Probability points y_m for mth extremes, m counted from above.

Table 5 Cumulative probability Ψ and density function ψ for the reduced range R

Table 6. Reduced range R as a function of the cumulative probability Ψ.

The first three tables will be found indispensable for work involving the application of the theory of extreme values. The fourth table gives the values of the reduced variate y_m at the probability points 005, 010, .025, 050, 100 and .500 at both ends of the distribution, for the 15 largest sample values and beyond this at intervals of 5 up to $m=50$. The distribution function is the asymptotic distribution of the doubly exponential type discussed on page 27 of Gumbel's published lectures.

Tables 5 and 6 are based on the asymptotic distribution of the "reduced range" R Although some explanation of the concept of reduced range is given in the Introduction, it might be well to have in mind the relation

$$R = \frac{\pi}{\sqrt{3}} \frac{w - \bar{w}}{\sigma_w} + 2\gamma, \qquad \begin{aligned} \gamma &= \text{Euler's constant,} \\ w &= \text{observed range,} \end{aligned}$$

and refer to Gumbel's article in the *Annals of Mathematical Statistics*, "The Distribution of the Range" (Vol. 18 (1947) pp 384–412)

The tables in other respects are clearly explained in the Introduction As regards accuracy one finds the statement on page 11. "The entries in all these tables are guaranteed to about a unit in the last place given, with the possible exception of the table of probability points for mth extremes, which is correct to within several units in the last place, provided that Thompson's table (of percentage points of the χ^2-distribution) is correct to within several units in the last place." Parenthesis is supplied by this reviewer

With prices of most publications what they are today, it is distinctly a privilege to be able to obtain publications of this calibre at the prices quoted for these

REFERENCES

[1] Epstein B , "Application of the theory of extreme values in fracture problems," *Journal of the American Statistical Association*, 43 (1948), 403–12.

[2] Gumbel, E. J , *Statistical Theory of Extreme Values and Some Practical Ap-*

plications, A Series of Lectures, National Bureau of Standards, Applied Mathematics Series, 33. Washington: U S. Government Printing Office, 1954. Pp. viii, 51. $0.40.

[3] Gumbel, E J, "The return period of flood flows," *Annals of Mathematical Statistics*, 12 (1941), 163–90.

[4] Gumbel, E J, "Probability-interpretation of the observed return-periods of floods," *Transactions of the American Geophysical Union of 1941*, 836–50.

[5] Gumbel, E J and Lieblein, J, "Some applications of extreme value methods," *The American Statistician*, 8, Dec. 1954, 14–17.

[6] Gumbel, E. J. and von Schelling, H., "The distribution of the number of exceedances," *Annals of Mathematical Statistics*, 21 (1950), 247–62.

[7] Kase, Shigeo, "A theoretical analysis of the distribution of tensile strength of vulcanized rubber," *Journal of Polymer Science*, 11 (1953), 425–31.

[8] Kimball, B F., "Sufficient statistical estimation functions for the parameters of the distribution of maximum values," *Annals of Mathematical Statistics*, 17 (1946), 299–309 *Correction* The plus sign before partial derivative in (5.2) should be a minus sign The negative sign before the exponent αu in (5 4) should be a plus sign.

[9] Kimball, B F, "An approximation to the sampling variance of an estimated maximum value of given frequency based on fit of doubly exponential distribution," *Annals of Mathematical Statistics*, 20 (1949), 110–13.

[10] Kimball, B. F., "Assignment of frequencies to a completely ordered set of sample data," *Transactions of the American Geophysical Union*, 27 (1946), 843–46, and *Discussion*, *ibid*. 28 (1947), 951–53

[11] Lieblein, Julius, "A new method of analyzing extreme-value data," National Bureau of Standards, Technical Note 3053 (1954), 88

[12] Sherman, B, "A random variable related to the spacing of sample values," *Annals of Mathematical Statistics*, 21 (1950), 339–61.

[13] "Probability tables for the analysis of extreme-value data," with Introduction by E J Gumbal, National Bureau of Standards, Applied Mathematics Series, 22 Washington: U S Government Printing Office, 1953. Pp iii, 32. $0 25.

ESTIMATE OF THE INTEGRATED NORMAL CURVE BY MINIMUM NORMIT CHI-SQUARE WITH PARTICULAR REFERENCE TO BIO-ASSAY

JOSEPH BERKSON
Mayo Clinic

THE integrated normal distribution function may be written

$$P_i = 1/\sqrt{2\pi}\int_{-\infty}^{\nu_i} e^{-x^2/2}dx = 1/\sqrt{2\pi}\int_{-\infty}^{\alpha+\beta x_i} e^{-x^2/2}\,dx \tag{1}$$

where P_i, the value of P at $x=x_i$, is the area under the normal frequency curve from $x=-\infty$ to $x=x_i$, μ is the mean of the frequency function, σ is the standard deviation, $\alpha=-(\mu/\sigma)$, $\beta=1/\sigma$, and ν_i is given by (2).

The straight-line transform of (1) is[1]

$$\text{Normit } P_i = \nu_i = \frac{x_i - \mu}{\sigma} = \alpha + \beta x_i; \tag{2}$$

therefore if the normit of P is plotted against x, the points will fall on a straight line with α as the intercept and β as the slope.

The integrated normal curve (1) has been advanced extensively in recent years for use in bio-assay with "quantal response" Here P_i represents the true probability of death (or some other all-or-none effect) at $x=x_i$ In this statistical model it is assumed that p_i, the observed proportion of individuals affected out of n_i exposed at x_i, can be considered a random variable binomially distributed around the true P_i, with variance $\sigma^2_{p_i}=P_iQ_i/n_i$.

For the estimation of the parameters α, β of (1), the method of maximum likelihood has been advocated, on the basis of the optimum asymptotic properties of the maximum likelihood estimate [10]. In this paper is presented another estimate, formally analogous with the minimum logit χ^2 estimate of the logistic function,[2] called the "minimum normit χ^2 estimate"

[1] "Normit" is intended as a diminutive for the "normal deviate" of Galton and Sheppard [17], the widely used "probit" is equal to the normit plus 5 The normit is used in the present development rather than the probit, because of the greater simplicity resulting from the use of the normit Employment of the probit would have necessitated doubling the size of the tables of this paper, involved larger numbers in the calculations and, because of the arbitrary constant 5, confused mathematical analysis [11] [24] If "probit analysis" is the appropriate name for maximum likelihood estimation of the integrated normal curve with the use of probits [10], the analogous name of "normit analysis" can be used for estimation by minimum normit χ^2, with the use of normits

[2] The development of the minimum logit χ^2 estimate of the logistic function, however, was somewhat differently motivated, and that estimate has the property of sufficiency, which so far as I know, is not possessed by estimates of the integrated normal curve

The minimum normit χ^2 estimate is defined by minimization of the following quantity, called the "normit χ^2," which is asymptotically distributed as χ^2:

$$\chi^2 \text{ (normit)} = \sum n_i \frac{z_i^2}{p_i q_i} (\nu_i - \hat{\nu}_i)^2 \tag{3}$$

where n_i is the number exposed at x_i, $p_i = 1 - q_i$ is the observed proportion affected at x_i, ν_i is the observed normit at x_i, $\hat{\nu}_i$ is the estimate of the normit, z_i is the ordinate of the unit normal curve at the point where its area is divided into p_i and q_i, and is given by

$$z_i = \frac{1}{\sqrt{2\pi}} e^{-\hat{\nu}_i^2/2}. \tag{4}$$

Since the minimum normit χ^2 estimate is definitively calculated without iteration, while the maximum likelihood estimate is obtainable only as the limit of successive iterative procedures, the minimum normit χ^2 estimate is much more easily determinable precisely than the maximum likelihood estimate [1] [2].

Recently Taylor [27] has demonstrated that the minimum normit χ^2 estimate is in the class of estimates R.B A N. of Neyman, and consequently has the same asymptotic properties as the maximum likelihood estimate. The minimum normit χ^2 estimate, as well as the maximum likelihood estimate, is therefore asymptotically efficient If an estimator is asymptotically efficient, it is widely accepted as implying that with large samples the variance is minimum and given by $E(\partial \ln\phi/\partial\theta)^2$, where ϕ is the probability of a sample and θ is the parameter estimated. Although in strictly rigid mathematics, the properties associated with asymptotic efficiency imply nothing *necessarily* for finite samples, the approximation for large samples is applicable in many specific cases Even where an implication for finite samples is valid, what size sample is "large" and how close the approximation is, must be independently determined In order to get some idea for the present case, the following "large sample" experiment was tried: Three equally spaced dosages x, corresponding respectively to true P's 0 3, 0 5, 0 7; $n=50$ at each dose, $\beta=0.524401$, considered known, $\alpha=0$ to be estimated. A stratified random sample of 600 was used [3] Both the maximum likelihood esti-

[3] For the method of sampling used in this and other experiments presented, see reference [3]. For the minimum normit χ^2 estimate the noniterative procedure of calculation described in the present text was used, for the maximum likelihood estimate, the interative method of probit analysis was used employing the Finney-Stevens tables [12] and also the method of Garwood as presented with auxiliary tables by Cornfield and Mantel [7], supplemented by the W.P A tables [21].

mate and the minimum normit χ^2 estimate of α are unbiased for this dosage arrangement The variance of the maximum likelihood estimate was determined as 0.011191, the variance of the minimum normit χ^2 estimate was determined as 0 011186 The value of the asymptotic variance is equal to the reciprocal of $\sum(nZ^2/PQ)$ where n is the number exposed at the several dosage values, 50 in the present example, $P=1-Q$ is the true P at the dose levels and Z is the ordinate of the normal distribution function corresponding to the integral P For the present case, the asymptotic variance $\sigma^2_{\hat{\alpha}}=0\ 011186$, so that the variance of the minimum normit χ^2 estimate was found to be equal to the asymptotic value to the precision of six decimal figures, and the variance of the maximum likelihood estimate was found to be slightly larger than the asymptotic value. The extreme closeness of the variance of the minimum normit χ^2 estimate to the asymptotic value is impressive, and the finding of a somewhat larger variance for the maximum likelihood estimate is in agreement with experiments to be described below, but the present point is that with $n=50$ at each of three dose levels, it appears that both estimates have attained their asymptotic distributions to a very close degree of approximation and we may reasonably expect that the situation is not much different for other similar arrangements of dosages and also with both parameters to be estimated, when the total number in the experiment $\geqq 150$.

The finding of a greater variance for the maximum likelihood estimate as compared with the minimum normit χ^2 estimate, in the experiment with "large" n, as well as a similar previous finding for the analogous minimum logit χ^2 estimate [3], suggested the desirability of comparing the two estimators for "small samples " A series of experiments were carried out similar to those performed to compare the minimum logit χ^2 estimate with the maximum likelihood estimate of the logistic function [3], but not so extensive, each experiment here being based on a stratified random sample of 600 Experiments were performed as for three equally spaced doses, 10 at each dose, at dosage values scaled -1, 0, $+1$, the respective values of P were taken as 0 3, 0 5, 0.7, which defined the parameters as $\alpha=0$, $\beta=0\ 524401$

The program was in two sections, (1) β considered known, only α to be estimated, and (2) α and β both to be estimated simultaneously Eight experiments were performed, one for estimate of α, one for estimate of α and β, at each of the following values for central P: 0 5, 0.6, 0 7, and 0 8 The results are summarized in Table 1. It is seen that in each of the experiments the error variance around the mean and the mean-square-error from the true value of the parameter are smaller for

the minimum normit χ^2 estimate than for the maximum likelihood estimate. This result is in agreement with the comparison of the maximum likelihood estimate with the minimum logit χ^2 estimate for the estimation of the logistic function reported previously [3], but the difference in favor of the χ^2 estimate is generally greater in the case of the logistic function than with the integrated normal curve.

Of noteworthy interest is a consideration of the variances in relation to I, Fisher's "amount of extractable information." For estimation of the single parameter α, the value of the asymptotic variance is also that of the Cramer-Rao [8, 26] lower bound in finite samples for a regular unbiased estimate, and is equal to $1/I$. It has been noted in the "large sample" experiment with dosages symmetrically disposed about $P=0.5$, that the variance of the minimum normit χ^2 estimate was found to be practically equal to the asymptotic variance, while the variance of the maximum likelihood estimate was slightly above it For the same experiment with small samples the variance of the minimum normit χ^2 estimate in relation to $1/I$ is less than with large samples, so that with small samples, *it is less than $1/I$* On the other hand the variance of the maximum likelihood estimate in relation to $1/I$ increases with decrease of sample size, so that it is greater than $1/I$ The position of the variance of the minimum normit χ^2 estimate below $1/I$ is maintained for other dispositions of the dosages where the estimate is biased, and even the mean-square-error is less than $1/I$, while the variance of the maximum likelihood estimate is everywhere above $1/I$. Here again relations are similar to those found with the analogous estimates of the logistic function [3].

CALCULATION

The normal equations for obtaining the estimates of α and β are

$$\sum n_i w_i(\nu_i - \hat{\nu}_i) = 0, \tag{5}$$

$$\sum n_i w_i x_i(\nu_i - \hat{\nu}_i) = 0, \tag{6}$$

where $w_i = z_i^2/p_i q_i$, ν_i is the observed normit at x_i, and $\hat{\nu}_i$ is the estimated value of the normit.

The evaluation of (5)(6) leads to a procedure that amounts simply to obtaining a least squares solution of the straight line

$$\hat{\nu}_i = a + bx_i \tag{7}$$

with $n_i w_i$ as weight of the observation ν_i The values of a and b, which

are the estimates respectively of α, β, are given by

$$b = \frac{\sum n_i w_i(\nu_i - \bar{\nu})(x_i - \bar{x})}{\sum n_i w_i (x_i - \bar{x})^2} = \frac{\sum n_i w_i \nu_i x_i - \dfrac{\sum n_i w_i \nu_i \sum n_i w_i x_i}{\sum n_i w_i}}{\sum n_i w_i x_i^2 - \dfrac{(\sum n_i w_i x_i)^2}{\sum n_i w_i}}, \tag{8}$$

$$a = \bar{\nu} - b\bar{x} = \frac{\sum n_i w_i \nu_i - b \sum n_i w_i x_i}{\sum n_i w_i}, \tag{9}$$

$$\bar{x} = \frac{\sum n_i w_i x_i}{\sum n_i w_i}, \qquad \bar{\nu} = \frac{\sum n_i w_i \nu_i}{\sum n_i w_i}.$$

The $E\,D\ 50 = \gamma$, the dosage value x which produces 50 per cent effect, is the value of x_i in (1) for which $P_i = 0\ 5$, and is given by

$$\gamma = -\alpha/\beta. \tag{10}$$

The estimate of γ, represented by x_{50}, is given by

$$x_{50} = -a/b \tag{11}$$

We may write the estimated normit linear equation as

$$\hat{\nu}_i = a + bx_i = a' + b(x_i - \bar{x}) \tag{12}$$

where $a' = \bar{\nu}$.

If x is measured as the logarithm of the dose D, then $x_{50} = \log D_{50}$, where x_{50} is the estimate of γ, the value of x corresponding to a 50 per cent response, and D_{50} is the estimate of the actual dose producing this response. Formulas for variances of the estimates of the parameters may be written as follows.[4]

$$s_{a'}^2 = \frac{1}{\sum nw}, \qquad s_b^2 = \frac{1}{\sum nw(x - \bar{x})^2},$$

$$s_a^2 = s_{a'}^2 + \bar{x}^2 s_b^2,$$

$$s_{x_{50}}^2 = \frac{1}{b^2}\left[s_{a'}^2 + s_b^2(x_{50} - \bar{x})^2\right], \qquad s_{D_{50}}^2 = s_{x_{50}}^2 \left(\frac{D_{50}}{\log e}\right)^2.$$

An example of calculations is given in Table 2.

[4] These are the asymptotic variances, with estimates obtained from observations replacing the true values of the parameters. For development of the asymptotic variances, see references [10] and [18] For a discussion of the limited usefulness of formulas for the standard errors of the estimates see references [1, 8a].

Tables are provided for facilitation of the calculations. The refinement to which the tables are carried was determined in relation to the objective that the estimate should finally be written with a number of decimals corresponding to two significant figures in the standard error of the estimate. Several typical examples were tried to determine the minimum number of figures that should be carried in the tables, in order to assure a solution which is arithmetically correct to the number of figures required by this rule. One cannot be sure that the best decision was made, it is possible that fewer figures in the tables would have sufficed, and on the other hand that examples may be met for which the refinement of the tables is insufficient for the required precision.

Table 3, giving the normit for argument p, was constructed with the use of values given in *The Kelley Statistical Tables* [20]. A convenient table similar to that given here is Table 1 of Pearson's Tables, Part I [22], in which, however, the normit is tabled only to four decimal places. *The Kelley Statistical Tables* list the argument to four decimal places, and the normit is given with eight decimal places. When the observed p is either zero or 100 per cent, a substitute working value is used equal respectively to $1/2n$ and $1-1/2n$.[5]

It will be observed in equations (8)(9), which are the formulas for the estimates, that they contain $w=z^2/pq$ and $w\nu$, each of these being a function of the observed proportion p. Table 4 gives both these quantities for argument p. This table was constructed from values given in Table II of Pearson's *Tables for Statisticians and Biometricians*, Part II [23].

When an equation has been "fitted" to observed data, it is good practice to calculate the values "predicted" by the equation to compare directly with the values observed. Table 5, which gives the area under the normal frequency curve, in a form convenient for the present situation, can be used for such computation, it was calculated with the use of values in the W.P.A. tables of the normal function [21]. Table 5 is used also for the direct calculation of the Pearson $\chi^2=\sum(o-e)^2/e$. Values for χ^2, to be used in tests of significance employing available χ^2 tables, can be obtained more easily by calculation of the normit $\chi^2 = \sum nw(\nu-\hat{\nu})^2$ as shown in Table 2.[6] The χ^2 tables consulted in ac-

[5] For treatment of the case of zero survivors, see reference [1].

[6] The normit χ^2 can be calculated directly from the formula

$$\chi^2 \text{ (normit)} = [\sum n_i w_i(\nu_i-\bar{\nu})(x-\bar{x})]^2/\sum n_i w_i(x_i-\bar{x})^2,$$

but this is inadvisable because, (1) it necessitates carrying an extra column ($nw\nu$) in the computations, not required for the estimates themselves, (2) it is very important, in considering the differences between the observed values and the "predicted," to note the particular observations that make the major contributions to the total χ^2. Frequently a single discrepant "stray" observation may result in a large total

complishing a test of significance represent the asymptotic distribution of χ^2, and neither the Pearson χ^2 nor the normit χ^2 calculated from finite samples follows this distribution exactly For both, the χ^2 distribution is approached asymptotically as the n_i's$\rightarrow\infty$; however, so far as the present author is aware, it is not known which χ^2, the Pearson χ^2 or the normit χ^2, for finite samples, approximates more closely the distribution of the tabled χ^2. Objectively, therefore, on the basis of available knowledge, there is no reason for insisting on the Pearson χ^2 rather than the normit χ^2. However, it would seem to be in the order of proper statistical comportment to use the conventional Pearson χ^2 until another formulation is shown to be preferable For this reason, and also because the Pearson χ^2 involves the calculation of the expected $\widehat{p}$, which is itself of immediate interest, it is recommended that the Pearson χ^2 be used.

NOTATION ON THE HISTORY OF NORMIT ANALYSIS

Gaddum [14], in his original (1933) presentation of the application of the integrated normal curve to bio-assay, utilized the same weights based on the observed relative frequencies as employed in this paper Indeed he has specifically advocated these weights in preference to those required by maximum likelihood, on the grounds that the estimates are much simpler to obtain in this way and that the results are not much different Here Gaddum, who in some circumstances has insisted on "logic" in what is advanced in these matters [16], has not been entirely logical At the same time that he has urged the use of the observed weights for simplicity he advanced his opinion that maximum likelihood yields the best estimate of the normit line [15]. It is pertinent to ask, if a simple solution is desired, even one which admittedly is statistically inferior, *why use any weights at all*,—that is, why not use unit weight for all observations? Such a system of weighting yields consistent estimates [25], and they too usually are very close to those of maximum likelihood.

I have made no special investigation regarding "priority" in the use of the observed weights z^2/pq Urban (1910) [28], cited by Gaddum, developed the method outlined here—except for the treatment of zero survivors, and with a different motivation—and provided a table of the weights z^2/pq as well as a table of normits.[7] The method of normit analysis of this paper may therefore be regarded as the method of Urban, modified

The method of Urban apparently was in wide use for many years and

χ^2, which if the formula is used and therefore the individual contributions to the χ^2 are not apparent, may lead to a conclusion that the data as a whole are "heterogeneous."

[7] In terms of the normal curve as it is formulated in the physical sciences.

was incorporated in at least one standard textbook [19]. Since about 1938 it has been generally replaced by what has developed into a formidable statistical discipline now widely known under the name of "probit analysis" [5, 6, 9, 10, 13].

The dispossession of the method of Urban was effected, not on the basis of an objective examination, mathematical or experimental, of the merits of the established procedure compared with those of the new candidate, but solely, it appears, on the authority of a cryptically expressed opinion of Sir Ronald A. Fisher [4]. The findings reported in the present study indicate that not only is the older method much easier but, judged on the basis of variance and mean square of the resulting estimate, it is better.

REFERENCES

[1] Berkson, Joseph, "A statistically precise and relatively simple method of estimating the bio-assay with quantal response, based on the logistic function," *Journal of the American Statistical Association*, 48 (1953), 565–99.

[2] Berkson, Joseph, Comments on "The analysis of variance with various binomial transformations," by Professor Sir Ronald Fisher, *Biometrics*, 10 (1954), 147–51.

[3] Berkson, Joseph, "Maximum likelihood and minimum χ^2 estimates of the logistic function," *Journal of the American Statistical Association*, in press

[4] Bliss, C. I., "The calculation of the dosage-mortality curve"; Fisher, R. A. Appendix "The case of zero survivors," *Annals of Applied Biology*, 22 (1935), 134–64, 164–67

[5] Bliss, C. I., "The determination of the dosage-mortality curve from small numbers," *Quarterly Journal of Pharmacy and Pharmacology*, 11 (1938), 192–216.

[6] Burn, J. H., Finney, D. J., and Goodwin, L. G., *Biological Standardization* (Second edition). London: Oxford University Press, 1950

[7] Cornfield, J., and Mantel, N., "Some new aspects of the application of maximum likelihood to the calculation of the dosage response curve," *Journal of the American Statistical Association*, 45 (1950), 181–210.

[8] Cramér, Harald, *Mathematical Methods of Statistics*. Princeton. Princeton University Press, 1946

[8a] Dews, P. B., and Berkson, J., "On the error of bio-assay with quantal response," in *Statistics and Mathematics in Biology*, Ames, Iowa: The Iowa State College Press, 1954.

[9] Emmens, C. W., *Principles of Biological Assay*, London: Chapman and Hall Ltd., 1948.

[10] Finney, D. J., *Probit Analysis, A Statistical Treatment of the Sigmoid Response Curve*, Cambridge: The University Press, 1947.

[11] Finney, D. J., "The principles of biological assay." *Supplement to the Journal of the Royal Statistical Society*, 9 (1947), 46–91. See discussions by Fieller and Irwin.

(*Continued on page 549*)

TABLE 1

SAMPLING ERRORS OF MAXIMUM LIKELIHOOD AND MINIMUM NORMIT χ^2 ESTIMATES

(1) Estimate of α, β being known

True P at dose			Bias		Variance		Mean-square-error		
Low	Mid	High	Maximum likelihood	Minimum normit χ^2	Maximum likelihood	Minimum normit χ^2	Maximum likelihood	Minimum normit χ^2	$1/I$
.3	5	7	− 0003*	− 0013*	0585	0538	0585	0538	0559
.393	6	782	0051	− 0095	0623	0560	0623	0561	0571
5	7	853	0156	− 0169	0680	.0592	.0682	0595	.0612
624	8	914	0265	− 0376	0898	0644	0905	0658	0706

(2) α and β to be estimated

True P at dose			Bias		Variance		Mean-square-error	
Low	Mid	High	Maximum likelihood	Minimum normit χ^2	Maximum likelihood	Minimum normit χ^2	Maximum likelihood	Minimum normit χ^2
Estimate of α								
3	5	7	0003*	− 0003*	0675	0581	0675	0581
393	6	782	− 0054	− 0101	0888	0825	0888	0826
.5	7	853	− 0126	− 0039	1502	1424	1504	1424
624	8	914	− 0189	0358	3109	2682	3113	.2695
Estimate of β								
3	5	.7	0449	0272	1060	0930	.1080	0937
393	6	782	0525	.0251	1157	0930	1185	0936
5	.7	853	0553	0010	1164	0815	1195	0815
624	8	914	− 0230	− 0518	1227	.0758	1232	0785

* The estimates for this disposition of dosages are unbiased, the indicated values are sampling errors.

TABLE 2

EXAMPLE OF CALCULATIONS (DATA FROM FINNEY [8] P. 26)

Calculation of a, b, x_{50}									Calculation of Normit χ^2			
Concentration, mg /liter D	D/D_1 D'	Log D' x	Deaths r	Total no. n	Proportion died $r/n=p$	nw*	$nw\nu$*	nwx	ν	$\hat{\nu}$	$(\nu-\hat{\nu})^2$	$nw(\nu-\hat{\nu})^2$
2 6	1 00	0 000	6	50	0 120	18 9450	−22 2650	0	−1.1750	−1 1345	0 0016	0 0303
3 8	1 46	0 164	16	48	0 333	28 5456	−12 3216	4 6815	−0 4316	−0 4473	0 0002	0 0057
5 1	1 96	0 292	24	46	0 522	29 2514	1 6146	8 5414	0 0552	0 0891	0 0011	0 0322
7 7	2 96	0 471	42	49	0 857	20 3840	21 7511	9 6009	1 0669	0 8392	0 0518	1 0559
10 2	3 92	0 593	44	50	0 880	18 9450	22 2650	11 2344	1 1750	1 3505	0 0308	0 5835
												1 7076 $\chi^2=1$ 7

* w and $w\nu$ from Table 4

$\Sigma\, nw = 116$ 0710 $\qquad \Sigma\, nwx = 34$ 0582 $\qquad \Sigma\, nw\nu = 11$ 0441

$$\bar{x} = \frac{\Sigma\, nwx}{\Sigma\, nw} = 0\ 2934 \qquad \bar{\nu} = \frac{\Sigma\, nw\nu}{\Sigma\, nw} = 0\ 09515$$

$$\Sigma\, nwx^2 = 14\ 4459 \qquad \Sigma\, nw\nu x = 21\ 8986$$

$$\frac{(\Sigma\, nwx)^2}{\Sigma\, nw} = 9\ 9935 \qquad \frac{\Sigma\, nw\nu\, \Sigma\, nwx}{\Sigma\, nw} = 3\ 2406$$

$$\Sigma\, nw(x-\bar{x})^2 = \text{Diff} = 4\ 4524 \qquad \Sigma\, nw(\nu-\bar{\nu})(x-\bar{x}) = \text{Diff} = 18\ 6580$$

$$b = \frac{\Sigma\, nw(\nu-\bar{\nu})(x-\bar{x})}{\Sigma\, nw(x-\bar{x})^2} = 4\ 1905 \qquad a = \frac{\Sigma\, nw\nu - b\,\Sigma\, nwx}{\Sigma\, nw} = -1\ 1345$$

$$x_{50} = -\frac{a}{b} = 0\ 2707 \qquad \log D_{50} = x_{50} + \log D_1 = 0\ 2707 + 0\ 4150 = 0\ 6857 \qquad D_{50} = 4\ 8495$$

$$s_{a'}^2 = \frac{1}{\Sigma\, nw} = 0.008615$$

$$s_b^2 = \frac{1}{\Sigma\, nw(x - \bar{x})^2} = 0.2246 \qquad s_b = 0.47$$

$$s_a^2 = s_{a'}^2 + \bar{x}^2 s_b^2 = 0.02795 \qquad s_a = 0.17$$

$$s_{x_{50}}^2 = s_{\log D_{50}}^2 = \frac{1}{b^2}\left[s_{a'}^2 + (x_{50} - \bar{x})^2 s_b^2\right] = 0.0004972 \qquad s_{x_{50}} = 0.022$$

$$s_{D_{50}}^2 = s_{\log D_{50}}^2 \left(\frac{D_{50}}{\log e}\right)^2 = 0.06199 \qquad s_{D_{50}} = 0.25$$

$$a = -1.13 \pm 0.17 \qquad b = 4.19 \pm 0.47† \qquad x_{50} = 0.271 \pm 0.022 \qquad D_{50} = 4.85 \pm 0.25†$$

$$\hat{\nu} = 4.1905x - 1.1345$$

DIRECT CALCULATION OF χ^2

x	n	p	$\hat{\nu}$	$\hat{p}$‡	$(p-\hat{p})^2$	$\widehat{pq}$	$\frac{n}{\hat{p}\hat{q}}$	$\frac{n}{\hat{p}\hat{q}}(p-\hat{p})^2$
0.000	50	0.120	−1.1345	0.12830	0.00007	0.11184	447.1	0.0313
0.164	48	0.333	−0.4473	0.32732	0.00003	0.22018	218.0	0.0065
0.292	46	0.522	0.0891	0.53550	0.00018	0.24874	184.9	0.0333
0.471	49	0.857	0.8392	0.79933	0.00333	0.16040	305.5	1.0173
0.593	50	0.880	1.3505	0.91157	0.00100	0.08061	620.3	0.6203
								1.7087

$\chi^2 = 1.71$

† The estimates given by Finney are $b = 4.01$, $D_{50} = 4.86$, the maximum likelihood estimates correct to two decimal places are $b = 4.22$, $D_{50} = 4.84$.
‡ Obtained from antinormit Table 5 by linear interpolation.

TABLE 3

NORMITS, ν

For p less than .50 on left, normit is negative. For p greater than .50 on right, normit is positive

p	Thousandths, for p in left column											
	0	1	2	3	4	5	6	7	8	9	—	
.00	∞	3.09023	2.87816	2.74778	2.65207	2.57583	2.51214	2.45726	2.40892	2.36562	2.32635	.99
.01	2.32635	2.29037	2.25713	2.22621	2.19729	2.17009	2.14441	2.12007	2.09693	2.07485	2.05375	.98
.02	2.05375	2.03352	2.01409	1.99539	1.97737	1.95996	1.94313	1.92684	1.91104	1.89570	1.88079	.97
.03	1.88079	1.86630	1.85218	1.83842	1.82501	1.81191	1.79912	1.78661	1.77438	1.76241	1.75069	.96
.04	1.75069	1.73920	1.72793	1.71689	1.70604	1.69540	1.68494	1.67466	1.66456	1.65463	1.64485	.95
.05	1.64485	1.63523	1.62576	1.61644	1.60725	1.59819	1.58927	1.58047	1.57179	1.56322	1.55477	.94
.06	1.55477	1.54643	1.53820	1.53007	1.52204	1.51410	1.50626	1.49851	1.49085	1.48328	1.47579	.93
.07	1.47579	1.46838	1.46106	1.45381	1.44663	1.43953	1.43250	1.42554	1.41865	1.41183	1.40507	.92
.08	1.40507	1.39838	1.39174	1.38517	1.37866	1.37220	1.36581	1.35946	1.35317	1.34694	1.34076	.91
.09	1.34076	1.33462	1.32854	1.32251	1.31652	1.31058	1.30469	1.29884	1.29303	1.28727	1.28155	.90
.10	1.28155	1.27587	1.27024	1.26464	1.25908	1.25357	1.24808	1.24264	1.23723	1.23186	1.22653	.89
.11	1.22653	1.22123	1.21596	1.21073	1.20553	1.20036	1.19522	1.19012	1.18504	1.18000	1.17499	.88
.12	1.17499	1.17000	1.16505	1.16012	1.15522	1.15035	1.14551	1.14069	1.13590	1.13113	1.12639	.87
.13	1.12639	1.12168	1.11699	1.11232	1.10768	1.10306	1.09847	1.09390	1.08935	1.08482	1.08032	.86
.14	1.08032	1.07584	1.07138	1.06694	1.06252	1.05812	1.05374	1.04939	1.04505	1.04073	1.03643	.85
.15	1.03643	1.03215	1.02789	1.02365	1.01943	1.01522	1.01103	1.00686	1.00271	0.99858	0.99446	.84
.16	0.99446	0.99036	0.98627	0.98220	0.97815	0.97411	0.97009	0.96609	0.96210	0.95812	0.95417	.83
.17	0.95417	0.95022	0.94629	0.94238	0.93848	0.93459	0.93072	0.92686	0.92301	0.91918	0.91537	.82
.18	0.91537	0.91156	0.90777	0.90399	0.90023	0.89647	0.89273	0.88901	0.88529	0.88159	0.87790	.81
.19	0.87790	0.87422	0.87055	0.86689	0.86325	0.85962	0.85600	0.85239	0.84879	0.84520	0.84162	.80
.20	0.84162	0.83805	0.83450	0.83095	0.82742	0.82389	0.82038	0.81687	0.81338	0.80990	0.80642	.79
.21	0.80642	0.80296	0.79950	0.79606	0.79262	0.78919	0.78577	0.78237	0.77897	0.77557	0.77219	.78
.22	0.77219	0.76882	0.76546	0.76210	0.75875	0.75542	0.75208	0.74876	0.74545	0.74214	0.73885	.77
.23	0.73885	0.73556	0.73228	0.72900	0.72574	0.72248	0.71923	0.71599	0.71275	0.70952	0.70630	.76
.24	0.70630	0.70309	0.69988	0.69668	0.69349	0.69031	0.68713	0.68396	0.68080	0.67764	0.67449	.75

p	Thousandths, for p in left column											
	0	1	2	3	4	5	6	7	8	9	—	
.25	0.67449	0.67135	0.66821	0.66508	0.66196	0.65884	0.65573	0.65262	0.64952	0.64643	0.64335	.74
.26	0.64335	0.64027	0.63719	0.63412	0.63106	0.62801	0.62496	0.62191	0.61887	0.61584	0.61281	.73
.27	0.61281	0.60979	0.60678	0.60376	0.60076	0.59776	0.59477	0.59178	0.58879	0.58581	0.58284	.72
.28	0.58284	0.57987	0.57691	0.57395	0.57100	0.56805	0.56511	0.56217	0.55924	0.55631	0.55338	.71
.29	0.55338	0.55047	0.54755	0.54464	0.54174	0.53884	0.53594	0.53305	0.53016	0.52728	0.52440	.70
.30	0.52440	0.52153	0.51866	0.51579	0.51293	0.51007	0.50722	0.50437	0.50153	0.49869	0.49585	.69
.31	0.49585	0.49302	0.49019	0.48736	0.48454	0.48173	0.47891	0.47610	0.47330	0.47050	0.46770	.68
.32	0.46770	0.46490	0.46211	0.45933	0.45654	0.45376	0.45099	0.44821	0.44544	0.44268	0.43991	.67
.33	0.43991	0.43715	0.43440	0.43164	0.42889	0.42615	0.42340	0.42066	0.41793	0.41519	0.41246	.66
.34	0.41246	0.40974	0.40701	0.40429	0.40157	0.39886	0.39614	0.39343	0.39073	0.38802	0.38532	.65
.35	0.38532	0.38262	0.37993	0.37723	0.37454	0.37186	0.36917	0.36649	0.36381	0.36113	0.35846	.64
.36	0.35846	0.35579	0.35312	0.35045	0.34779	0.34513	0.34247	0.33981	0.33716	0.33450	0.33185	.63
.37	0.33185	0.32921	0.32656	0.32392	0.32128	0.31864	0.31600	0.31337	0.31074	0.30811	0.30548	.62
.38	0.30548	0.30286	0.30023	0.29761	0.29499	0.29237	0.28976	0.28715	0.28454	0.28193	0.27932	.61
.39	0.27932	0.27671	0.27411	0.27151	0.26891	0.26631	0.26371	0.26112	0.25853	0.25594	0.25335	.60
.40	0.25335	0.25076	0.24817	0.24559	0.24301	0.24043	0.23785	0.23527	0.23269	0.23012	0.22754	.59
.41	0.22754	0.22497	0.22240	0.21983	0.21727	0.21470	0.21214	0.20957	0.20701	0.20445	0.20189	.58
.42	0.20189	0.19934	0.19678	0.19422	0.19167	0.18912	0.18657	0.18402	0.18147	0.17892	0.17637	.57
.43	0.17637	0.17383	0.17128	0.16874	0.16620	0.16366	0.16112	0.15858	0.15604	0.15351	0.15097	.56
.44	0.15097	0.14843	0.14590	0.14337	0.14084	0.13830	0.13577	0.13324	0.13072	0.12819	0.12566	.55
.45	0.12566	0.12314	0.12061	0.11809	0.11556	0.11304	0.11052	0.10799	0.10547	0.10295	0.10043	.54
.46	0.10043	0.09791	0.09540	0.09288	0.09036	0.08784	0.08533	0.08281	0.08030	0.07778	0.07527	.53
.47	0.07527	0.07276	0.07024	0.06773	0.06522	0.06271	0.06020	0.05768	0.05517	0.05266	0.05015	.52
.48	0.05015	0.04764	0.04513	0.04263	0.04012	0.03761	0.03510	0.03259	0.03008	0.02758	0.02507	.51
.49	0.02507	0.02256	0.02005	0.01755	0.01504	0.01253	0.01003	0.00752	0.00501	0.00251	0.00000	.50
	—	9	8	7	6	5	4	3	2	1	0	p
	Thousandths, for p in right column											

TABLE 4

NORMIT WEIGHTS

Upper figure in table is $w = z^2/pq$; lower figure is wv

For p less than .50 on left, wv is negative. For p greater than .50 on right, wv is positive

p	Thousandths, for p on left											
	0	1	2	3	4	5	6	7	8	9		
.00		0113	0201	0280	0352	.0420	.0485	.0546	0605	0662	.0718	99
		0351	0580	0769	0934	.1083	1218	1342	1458	1567	1669	
01	0718	0771	0823	0873	0923	.0971	1018	1064	1109	1153	1196	.98
	1669	1766	1857	1944	2027	.2106	2182	2255	2325	2392	2456	
.02	1196	1239	.1280	1321	1362	1401	.1440	1479	1517	.1554	1591	97
	2456	2519	.2579	2637	2693	2747	2799	2850	2899	2946	.2992	
03	1591	1627	.1663	1698	1733	1768	.1802	1835	1869	1901	.1934	96
	2992	3037	.3080	3122	3163	3203	.3242	3279	3316	.3351	.3386	
04	.1934	.1966	1998	2029	2060	2091	2121	2151	2181	.2210	2239	95
	3386	.3419	3452	3483	3514	.3544	3574	3602	3630	.3657	3683	
05	2239	2268	2297	2325	2353	.2381	2408	2436	2463	.2489	2516	.94
	3683	.3709	.3734	3758	3782	.3805	3828	3850	3871	.3892	3912	
06	2516	.2542	.2568	2594	2620	.2645	2670	2695	2720	.2745	2769	.93
	3912	.3932	3951	3969	3988	4005	4022	4039	4056	4071	4087	
07	2769	2793	2817	2841	2865	.2888	2912	2935	2958	.2980	3003	92
	4087	4102	.4116	4131	4144	4158	.4171	4184	4196	4208	.4219	
08	3003	3025	.3048	3070	3092	3113	3135	3156	3178	.3199	3220	91
	4219	4231	4241	4252	4262	.4272	.4282	4291	4300	4309	4317	
09	3220	3241	3261	3282	3302	3323	.3343	3363	3383	3403	3422	90
	4317	4325	.4333	4340	4348	4355	4361	4368	4374	.4380	4386	
10	3422	3442	.3461	3480	3499	3518	.3537	.3556	3575	3593	.3612	89
	4386	4391	4396	4401	4406	4411	4415	4419	4423	4426	4430	
11	3612	3630	3648	3666	3684	3702	3720	.3737	3755	3772	3789	88
	4430	.4433	4436	4439	4441	4444	4446	4448	4450	4451	4453	
12	3789	3807	3824	3841	3858	.3874	3891	3908	3924	3940	3957	87
	4453	.4454	4455	4456	4456	4457	4457	4457	4457	4457	.4457	
13	3957	.3973	3989	4005	4021	4037	4053	4068	4084	4099	4115	86
	4457	4456	4456	4455	4454	4453	4452	.4450	4449	4447	4445	
14	4115	4130	4145	4160	4175	4190	4205	.4220	4235	4249	.4264	85
	4445	4443	4441	4439	4436	4434	4431	4428	4425	4422	4419	
15	.4264	4278	4293	4307	4321	4335	4349	4363	.4377	4391	4405	84
	4419	4416	.4412	.4409	4405	4401	4397	4393	4389	4385	4380	
16	4405	4418	4432	.4446	4459	4472	.4486	4499	4512	4525	4538	83
	4380	.4376	4371	4367	4362	4357	4352	4346	4341	4336	4330	
17	4538	.4551	4564	4577	4590	.4602	4615	4628	4640	4653	4665	82
	4330	4325	.4319	4313	4307	4301	4295	4289	4283	.4277	4270	
18	4665	4677	4689	4702	4714	.4726	4738	.4750	.4761	.4773	4785	81
	4270	4264	.4257	4250	4243	4236	4229	4222	4215	.4208	.4201	
19	4785	4797	4808	4820	4831	4843	.4854	4865	.4876	4888	4899	80
	.4201	4193	4186	4178	.4170	4163	.4155	.4147	4139	4131	4123	
20	.4899	4910	4921	4932	.4943	4953	4964	.4975	4985	4996	.5007	79
	.4123	4115	4106	4098	4090	4081	4072	.4064	4055	.4046	4037	
21	.5007	5017	5028	5038	.5048	5059	.5069	5079	5089	5099	5109	78
	.4037	4029	4020	4010	4001	3992	3983	.3974	3964	3955	3945	
.22	5109	5119	5129	5139	5149	5158	5168	.5178	5187	5197	5206	77
	3945	3936	3926	3916	.3906	3897	.3887	.3877	.3867	.3857	3847	
.23	5206	5216	5225	.5234	.5244	5253	.5262	5271	5280	5289	5298	76
	3847	3836	3826	.3816	.3806	3795	.3785	3774	3764	3753	3742	
.24	5298	5307	5316	5325	.5334	5343	.5351	5360	5369	5377	5386	75
	.3742	3732	.3721	3710	3699	.3688	3677	3666	3655	3644	3633	
		9	8	7	6	5	4	3	2	1	0	p
		Thousandths, for p on right										

TABLE 4—*Continued*

NORMIT WEIGHTS

Upper figure in table is $w = z^2/pq$; lower figure is w_v

For p less than .50 on left, w_v is negative For p greater than .50 on right, w_v is positive

p	Thousandths, for p on left											
	0	1	2	3	4	5	6	7	8	9		
25	.5386	5394	.5403	5411	5419	.5428	5436	5444	5452	5460	.5468	74
	3633	.3621	3610	.3599	3587	3576	3564	3553	3541	3530	3518	
26	5468	5476	5484	5492	5500	5508	.5516	5524	5531	.5539	5547	73
	3518	3506	.3495	3483	3471	3459	3447	3435	3423	3411	3399	
27	5547	5554	.5562	5569	5577	5584	5592	5599	5606	5614	5621	72
	3399	3387	3375	3363	3350	3338	3326	3313	3301	.3289	3276	
28	.5621	5628	5635	5642	5649	5656	5663	5670	5677	.5684	.5691	71
	.3276	3264	3251	3238	3226	3213	3200	.3188	3175	3162	3149	
29	5691	5697	5704	5711	.5718	5724	5731	.5737	5744	5750	5757	70
	.3149	.3136	3123	3110	3097	3084	3071	3058	3045	.3032	3019	
30	.5757	5763	5769	5776	5782	5788	5794	5801	5807	5813	5819	69
	3019	3006	.2992	2979	2966	2952	2939	2926	2912	2899	2885	
31	5819	5825	5831	5837	5843	5848	5854	5860	5866	5871	5877	68
	2885	2872	2858	2845	2831	2817	2804	2790	2776	2762	2749	
.32	5877	5883	5888	5894	5899	5905	5910	5916	5921	5926	5932	67
	2749	2735	.2721	2707	2693	2679	2665	2652	2638	2624	2609	
33	5932	5937	5942	5947	5953	5958	5963	5968	5973	5978	5983	66
	2609	2595	2581	2567	2553	2539	2525	2511	2496	2482	2468	
34	5983	5988	5993	5998	6002	6007	6012	6017	6021	6026	6031	65
	.2468	2453	2439	.2425	2410	2396	2382	2367	.2353	2338	.2324	
35	6031	6035	6040	6044	6049	6053	6058	6062	6066	6071	6075	64
	2324	.2309	2295	2280	2265	2251	2236	2222	2207	2192	2178	
36	6075	6079	6083	6087	6092	6096	6100	6104	6108	6112	6116	63
	2178	2163	2148	2133	2119	2104	2089	2074	2059	2044	2030	
37	6116	6120	6124	6127	6131	6135	6139	6142	6146	6150	6153	62
	2030	2015	2000	1985	1970	1955	1940	1925	1910	1895	1880	
38	6153	6157	.6161	6164	6168	6171	6174	6178	6181	6185	6188	61
	1880	1865	1850	1835	1819	1804	1789	1774	1759	1744	1728	
39	6188	6191	6194	.6198	6201	6204	6207	6210	6213	6216	6219	60
	1728	1713	1698	1683	1667	1652	1637	1622	1606	1591	1576	
40	6219	.6222	6225	6228	6231	6234	6236	6239	6242	6245	6247	59
	1576	1560	.1545	1530	1514	1499	1483	1468	1452	1437	1422	
41	6247	6250	.6253	6255	6258	6260	6263	.6265	6268	6270	6272	58
	1422	1406	1391	1375	.1360	1344	1329	1313	1297	1282	1266	
42	6272	6275	6277	6279	6282	6284	6286	6288	6290	6293	6295	57
	1266	1251	1235	.1220	1204	.1188	1173	1157	1142	1126	1110	
43	6295	6297	6299	6301	6303	6304	.6306	6308	6310	6312	.6314	56
	1110	.1095	1079	1063	1047	.1032	1016	1000	0985	0969	0953	
44	6314	6315	6317	6319	6320	6322	6324	6325	6327	6328	6330	55
	0953	0937	0922	0906	0890	0874	0859	0843	.0827	0811	0795	
45	6330	6331	6333	6334	6335	6337	6338	6339	6341	.6342	6343	54
	0795	0780	0764	.0748	.0732	.0716	0700	0685	0669	0653	0637	
46	6343	6344	6345	6346	6347	6348	6349	.6350	6351	6352	6353	53
	0637	0621	0605	0589	0574	0558	.0542	0526	0510	.0494	0478	
.47	.6353	6354	6355	6356	6356	6357	6358	6359	6359	6360	.6360	52
	.0478	0462	0446	0430	0415	0399	0383	0367	0351	0334	0319	
48	.6360	6361	6361	6362	6362	6363	6363	6364	6364	6364	6365	51
	.0319	0303	.0287	0271	0255	0239	0223	0207	.0191	0176	0160	
.49	.6365	6365	.6365	6365	6366	.6366	.6366	.6366	.6366	6366	.6366	50
	0160	.0144	0128	.0112	.0096	0080	0064	0048	0032	0016	0000	
		9	8	7	6	5	4	3	2	1	0	p
		Thousandths, for p on right										

TABLE 5

ANTINORMITS*

ν	Thousandths 0	1	2	3	4	5	6	7	8	9	ν
.00	.50000	50040	50080	50120	50160	50199	50239	50279	50319	50359	.00
.01	50399	50439	50479	50519	50559	50598	50638	50678	50718	50758	.01
.02	50798	50838	50878	50918	50957	50997	51037	51077	51117	51157	.02
.03	51197	51237	51276	51316	51356	51396	51436	51476	51516	51556	.03
.04	51595	51635	51675	51715	51755	51795	51835	51874	51914	51954	.04
.05	51994	52034	52074	52113	52153	52193	52233	52273	52313	52352	.05
.06	52392	52432	52472	52512	52552	52591	52631	52671	52711	52751	.06
.07	52790	52830	52870	52910	52950	52989	53029	53069	53109	53148	.07
.08	53188	53228	53268	53307	53347	53387	53427	53466	53506	53546	.08
.09	53586	53625	53665	53705	53745	53784	53824	53864	53903	53943	.09
.10	53983	54023	54062	54102	54142	54181	54221	54261	54300	54340	.10
.11	54380	54419	54459	54498	54538	54578	54617	54657	54696	54736	.11
.12	54776	54815	54855	54895	54934	54974	55013	55053	55093	55132	.12
.13	55172	55211	55251	55290	55330	55369	55409	55448	55488	55528	.13
.14	55567	55607	55646	55686	55725	55764	55804	55843	55883	55922	.14
.15	55962	56001	56041	56080	56120	56159	56198	56238	56277	56317	.15
.16	56356	56395	56435	56474	56513	56553	56592	56632	56671	56710	.16
.17	56750	56789	56828	56867	56907	56946	56985	57025	57064	57103	.17
.18	57142	57182	57221	57260	57299	57339	57378	57417	57456	57495	.18
.19	57535	57574	57613	57652	57691	57730	57770	57809	57848	57887	.19
.20	57926	57965	58004	58043	58082	58121	58160	58200	58239	58278	.20
.21	58317	58356	58395	58434	58473	58512	58551	58590	58629	58668	.21
.22	58706	58745	58784	58823	58862	58901	58940	58979	59018	59057	.22
.23	59095	59134	59173	59212	59251	59290	59328	59367	59406	59445	.23
.24	59484	59522	59561	59600	59638	59677	59716	59755	59793	59832	.24
.25	59871	59909	59948	59987	60025	60064	60102	60141	60180	60218	.25
.26	60257	60295	60334	60372	60411	60450	60488	60527	60565	60604	.26
.27	60642	60680	60719	60757	60796	60834	60873	60911	60949	60988	.27
.28	61026	61065	61103	61141	61180	61218	61256	61294	61333	61371	.28
.29	61409	61447	61486	61524	61562	61600	61639	61677	61715	61753	.29
.30	61791	61829	61867	61905	61944	61982	62020	62058	62096	62134	.30
.31	62172	62210	62248	62286	62324	62362	62400	62438	62476	62514	.31
.32	62552	62590	62627	62665	62703	62741	62779	62817	62854	62892	.32
.33	62930	62968	63006	63043	63081	63119	63156	63194	63232	63270	.33
.34	63307	63345	63382	63420	63458	63495	63533	63570	63608	63646	.34
.35	63683	63721	63758	63796	63833	63871	63908	63945	63983	64020	.35
.36	64058	64095	64132	64170	64207	64244	64282	64319	64356	64394	.36
.37	64431	64468	64505	64543	64580	64617	64654	64691	64728	64766	.37
.38	64803	64840	64877	64914	64951	64988	65025	65062	65099	65136	.38
.39	65173	65210	65247	65284	65321	65358	65395	65432	65469	65505	.39
.40	65542	65579	65616	65653	65689	65726	65763	65800	65836	65873	.40
.41	65910	65946	65983	66020	66056	66093	66130	66166	66203	66239	.41
.42	66276	66312	66349	66385	66422	66458	66495	66531	66567	66604	.42
.43	66640	66677	66713	66749	66786	66822	66858	66894	66931	66967	.43
.44	67003	67039	67076	67112	67148	67184	67220	67256	67292	67328	.44
.45	67365	67401	67437	67473	67509	67545	67581	67616	67652	67688	.45
.46	67724	67760	67796	67832	67868	67903	67939	67975	68011	68047	.46
.47	68082	68118	68154	68189	68225	68261	68296	68332	68368	68403	.47
.48	68439	68474	68510	68545	68581	68616	68652	68687	68723	68758	.48
.49	68793	68829	68864	68899	68935	68970	69005	69041	69076	69111	.49
	0	1	2	3	4	5	6	7	8	9	

* The table gives the value of p for a specified value of the normit ν. If ν is negative, p is 1 minus the tabled value.

TABLE 5—*Continued*

ANTINORMITS

p	0	1	2	3	4	5	6	7	8	9	p
	Thousandths										
.50	.69146	.69181	69217	69252	69287	69322	69357	.69392	.69427	69462	50
51	69497	69532	69567	69602	69637	69672	69707	.69742	69777	69812	51
.52	69847	69882	69917	69951	69986	70021	70056	.70090	70125	70160	52
53	70194	70229	70264	.70298	.70333	70368	70402	.70437	70471	70506	53
54	70540	70575	70609	70644	70678	70712	70747	.70781	70815	70850	54
55	.70884	70918	70953	70987	.71021	.71055	.71089	71124	.71158	71192	.55
56	71226	71260	.71294	71328	71362	71396	71430	.71464	71498	.71532	.56
57	71566	71600	71634	71668	71702	71735	71769	.71803	71837	.71871	57
.58	71904	71938	71972	72005	72039	72073	72106	72140	.72173	.72207	.58
59	72240	72274	72307	.72341	.72374	72408	72441	72475	72508	72541	59
60	72575	72608	72641	72675	.72708	.72741	72774	72807	.72841	72874	60
61	72907	.72940	72973	73006	73039	73072	73105	73138	73171	73204	61
62	73237	73270	73303	73336	73369	.73401	73434	73467	73500	73533	62
.63	.73565	73598	.73631	.73663	73696	.73729	.73761	73794	73826	73859	63
64	73891	73924	73956	73989	.74021	74054	.74086	74118	74151	74183	64
65	74215	74248	74280	74312	74344	.74377	74409	.74441	74473	74505	.65
66	.74537	74569	.74601	74633	74666	74698	74729	.74761	74793	74825	66
67	74857	74889	74921	.74953	74984	75016	75048	75080	.75111	75143	67
68	75175	75206	75238	75270	.75301	75333	75364	75396	.75427	75459	68
69	75490	75522	.75553	75585	75616	75647	.75679	75710	75741	.75772	69
70	75804	75835	75866	.75897	75928	75960	.75991	.76022	76053	76084	70
71	76115	76146	76177	76208	76239	76270	76300	.76331	.76362	76393	71
72	76424	76455	76485	76516	76547	76577	76608	.76639	.76669	76700	72
73	76731	76761	.76792	.76822	76853	76883	76913	.76944	.76974	77005	73
74	77035	77065	77096	77126	77156	77186	77217	.77247	77277	77307	74
.75	.77337	77367	.77397	77428	77458	77488	.77518	.77548	77577	.77607	75
76	.77637	77667	77697	77727	77757	77786	77816	77846	77876	.77905	76
77	77935	77965	77994	78024	78053	78083	78113	78142	78172	78201	77
.78	78230	78260	78289	78319	.78348	78377	.78407	78436	78465	.78494	.78
.79	78524	78553	78582	78611	78640	78669	.78698	.78727	78757	78786	79
80	78814	78843	78872	78901	78930	.78959	78988	.79017	79045	79074	80
81	79103	79132	79160	79189	79218	79246	79275	79304	.79332	79361	81
82	.79389	79418	79446	79475	79503	79531	79560	79588	.79617	79645	82
83	79673	79701	79730	79758	79786	79814	.79842	.79870	.79898	79927	.83
84	79955	79983	80011	80039	80067	80094	80122	80150	80178	80206	84
85	80234	80262	80289	80317	.80345	80372	80400	80428	80455	80483	85
86	80511	80538	80566	80593	80621	80648	80676	80703	80730	80758	86
87	80785	80812	80840	80867	80894	80921	80949	80976	.81003	.81030	87
88	81057	81084	81111	81138	81165	81192	81219	81246	81273	81300	88
89	81327	81354	81380	81407	81434	81461	81487	81514	81541	81567	89
90	81594	81621	81647	81674	81700	81727	81753	81780	81806	81833	90
91	81859	81885	81912	81938	81964	81990	82017	82043	82069	82095	91
92	82121	82148	82174	82200	82226	82252	82278	82304	82330	82356	92
93	82381	82407	82433	82459	82485	82511	.82536	82562	82588	82613	93
94	82639	82665	82690	82716	82742	82767	82793	82818	82844	82869	.94
95	.82894	82920	82945	82971	82996	83021	83046	83072	83097	83122	95
.96	83147	.83172	83198	83223	83248	83273	83298	83323	83348	83373	.96
97	83398	.83423	.83447	83472	83497	83522	83547	83572	83596	83621	.97
98	83646	.83670	83695	83720	83744	83769	83793	83818	83842	83867	.98
.99	83891	.83916	.83940	83965	83989	84013	.84038	84062	84086	84110	99
	0	1	2	3	4	5	6	7	8	9	

TABLE 5—*Continued*

ANTINORMITS

ν	Thousandths 0	1	2	3	4	5	6	7	8	9	ν
1 00	84134	84159	84183	84207	84231	84255	84279	84303	.84327	84351	1 00
1 01	84375	84399	84423	84447	84471	84495	84519	84542	84566	84590	1 01
1 02	84614	84637	84661	84685	84708	84732	84755	84779	84802	84826	1 02
1 03	84850	84873	84896	84920	84943	84967	84990	85013	85037	85060	1 03
1 04	85083	85106	85129	85153	85176	85199	85222	85245	85268	85291	1 04
1 05	85314	85337	85360	85383	85406	85429	85452	85474	85497	85520	1 05
1 06	85543	85566	85588	85611	85634	85656	85679	85701	85724	85747	1 06
1 07	85769	85792	85814	85836	85859	85881	.85904	85926	.85948	85971	1 07
1 08	85993	86015	86037	86060	86082	86104	86126	86148	86170	86192	1 08
1 09	86214	86236	86258	86280	86302	.86324	86346	86368	86390	86412	1 09
1 10	86433	86455	86477	86499	86520	86542	86564	86585	86607	86629	1 10
1 11	86650	86672	86693	86715	86736	86758	86779	86800	86822	86843	1 11
1 12	86864	86886	86907	86928	86949	86971	86992	87013	87034	87055	1 12
1 13	87076	87097	87118	87139	87160	87181	87202	87223	87244	87265	1 13
1 14	87286	87307	87327	87348	87369	87390	87410	87431	87452	87472	1 14
1 15	87493	87513	87534	87555	87575	87596	87616	87636	87657	87677	1 15
1 16	87698	87718	87738	87759	87779	87799	87819	87840	87860	87880	1 16
1 17	87900	87920	87940	87960	87980	88000	88020	88040	88060	88080	1 17
1 18	88100	88120	88140	88160	88179	88199	88219	88239	.88258	88278	1 18
1 19	88298	88317	88337	88357	88376	88396	88415	88435	88454	88474	1 19
1 20	88493	88512	88532	88551	88571	88590	88609	88628	88648	88667	1 20
1 21	88686	88705	88724	88744	88763	88782	88801	88820	88839	88858	1 21
1 22	.88877	.88896	.88915	.88934	88952	88971	.88990	89009	89028	.89046	1 22
1 23	89065	89084	89103	89121	89140	89158	89177	89196	89214	89233	1 23
1 24	89251	89270	89288	89307	89325	89343	89362	89380	89398	89417	1 24
1 25	89435	89453	89472	89490	89508	89526	89544	89562	89580	89599	1 25
1 26	89617	89635	89653	89671	89689	89706	89724	89742	89760	89778	1 26
1 27	89796	89814	89831	89849	89867	89885	89902	89920	89938	89955	1 27
1 28	89973	89990	90008	90025	90043	90060	90078	90095	90113	90130	1 28
1 29	90147	90165	90182	90199	90217	90234	90251	90268	90286	90303	1 29
1 30	90320	.90337	90354	90371	90388	90405	90422	90439	90456	90473	1 30
1 31	90490	90507	90524	90541	90558	90575	90591	90608	90625	90642	1 31
1 32	90658	90675	90692	90708	90725	90741	90758	90775	90791	90808	1 32
1 33	90824	90841	90857	90873	90890	90906	90923	90939	90955	90971	1 33
1 34	90988	91004	91020	91036	91053	91069	.91085	91101	91117	91133	1 34
1 35	91149	91165	91181	91197	91213	91229	91245	91261	91277	91293	1 35
1 36	91309	91324	91340	91356	91372	91387	91403	91419	91434	91450	1 36
1 37	91466	91481	91497	91512	91528	91543	91559	91574	91590	91605	1 37
1 38	91621	91636	91651	91667	91682	91697	91713	91728	91743	91758	1 38
1 39	91774	91789	91804	.91819	91834	91849	91864	91879	91894	91909	1 39
1 40	91924	91939	91954	91969	91984	91999	92014	92029	92043	92058	1 40
1.41	92073	92088	92103	92117	92132	.92147	92161	92176	92190	92205	1 41
1 42	92220	92234	92249	92263	92278	.92292	92307	92321	92335	92350	1 42
1 43	92364	92379	92393	92407	92421	92436	92450	.92464	92478	92492	1 43
1 44	92507	92521	92535	92549	92563	92577	92591	.92605	92619	92633	1 44
1 45	.92647	92661	92675	92689	92703	92717	92730	.92744	.92758	92772	1 45
1 46	92786	92799	92813	92827	92840	92854	92868	92881	92895	92908	1 46
1 47	92922	92935	92949	92962	.92976	92989	93003	93016	.93030	93043	1 47
1 48	.93056	93070	93083	93096	93110	93123	93136	.93149	.93162	93176	1 48
1 49	93189	93202	93215	93228	.93241	93254	93267	.93280	93293	93306	1 49
	0	1	2	3	4	5	6	7	8	9	

TABLE 5—*Continued*

ANTINORMITS

p	Thousandths 0	1	2	3	4	5	6	7	8	9	*p*
1 50	93319	93332	.93345	93358	.93371	93384	93397	93409	93422	93435	1 50
1 51	93448	93461	93473	93486	93499	.93511	93524	.93537	93549	.93562	1 51
1.52	93574	93587	.93600	93612	93625	93637	93650	93662	93674	93687	1 52
1 53	93699	93712	93724	93736	93749	93761	93773	93785	93798	93810	1 53
1 54	93822	93834	93846	93858	93871	93883	93895	93907	93919	93931	1 54
1 55	93943	93955	93967	93979	93991	94003	94015	94027	.94038	94050	1 55
1 56	94062	94074	94086	94097	94109	94121	94133	94144	94156	94168	1 56
1.57	.94179	94191	94202	94214	94226	.94237	.94249	94260	94272	94283	1 57
1 58	94295	94306	94318	94329	94340	94352	94363	94374	94386	94397	1 58
1 59	94408	.94420	.94431	94442	94453	94464	94476	94487	94498	94509	1 59
1 60	94520	94531	94542	94553	94564	94575	94586	94597	94608	94619	1 60
1 61	94630	94641	94652	94663	94674	94684	94695	94706	94717	94728	1 61
1 62	94738	94749	94760	94771	94781	94792	94803	94813	94824	94834	1 62
1 63	94845	94856	94866	94877	94887	94898	94908	94919	94929	94939	1 63
1 64	94950	94960	.94971	94981	94991	95002	95012	95022	95032	95043	1 64
1 65	95053	95063	95073	95083	95094	95104	95114	95124	95134	95144	1 65
1 66	95154	95164	95174	95184	95194	95204	95214	95224	95234	95244	1 66
1 67	95254	95264	95274	95284	95293	95303	95313	95323	95333	95342	1 67
1 68	95352	95362	95372	95381	95391	95401	95410	95420	95429	95439	1 68
1 69	95449	95458	95468	95477	95487	95496	95506	95515	95525	95534	1 69
1 70	95543	95553	95562	95572	95581	95590	95600	95609	.95618	95627	1 70
1 71	95637	95646	95655	95664	.95674	95683	95692	95701	95710	95719	1 71
1 72	95728	95737	95747	95756	95765	95774	95783	95792	95801	95810	1 72
1 73	95819	95827	95836	95845	95854	95863	95872	95881	95889	95898	1 73
1 74	.95907	.95916	.95925	.95933	95942	95951	95959	.95968	95977	95985	1 74
1 75	95994	96003	96011	96020	96028	96037	96046	96054	96063	96071	1 75
1 76	96080	96088	96097	96105	96113	96122	96130	96139	96147	96155	1 76
1 77	96164	96172	96180	96189	96197	96205	96213	96222	96230	96238	1 77
1 78	96246	96254	96263	96271	96279	96287	96295	96303	96311	96319	1 78
1 79	96327	96335	96343	96351	96359	96367	96375	96383	96391	96399	1 79
1 80	96407	96415	96423	96431	96438	96446	96454	96462	96470	96477	1 80
1 81	96485	96493	96501	96508	96516	96524	96532	96539	96547	96554	1 81
1 82	96562	96570	96577	96585	96592	96600	96608	96615	96623	96630	1 82
1 83	96638	96645	96652	96660	96667	96675	96682	96690	96697	96704	1 83
1 84	96712	96719	96726	96734	96741	96748	96755	96763	96770	96777	1 84
1 85	96784	96792	96799	96806	96813	96820	96827	96834	96842	96849	1 85
1 86	96856	96863	96870	96877	96884	96891	96898	96905	96912	96919	1 86
1 87	96926	96933	96940	96947	96954	96960	96967	96974	96981	96988	1 87
1 88	96995	97001	97008	97015	97022	97029	97035	97042	97049	97055	1 88
1 89	97062	.97069	.97075	97082	97089	97095	97102	97109	97115	97122	1 89
1 90	97128	.97135	97141	97148	97155	97161	.97168	.97174	97180	97187	1 90
1 91	97193	97200	97206	97213	97219	97225	97232	97238	97244	97251	1 91
1 92	97257	97263	97270	97276	97282	97289	97295	97301	97307	97313	1 92
1 93	97320	.97326	97332	97338	97344	97351	97357	.97363	97369	97375	1 93
1 94	97381	97387	97393	97399	97405	97411	97417	.97423	97429	97435	1 94
1 95	97441	97447	97453	97459	97465	97471	97477	97483	97489	97494	1 95
1 96	97500	97506	97512	97518	97524	97529	97535	97541	97547	97552	1 96
1 97	.97558	.97564	97570	97575	97581	97587	97592	.97598	97604	97609	1 97
1 98	97615	97620	97626	.97632	.97637	97643	97648	97654	.97659	97665	1 98
1 99	.97670	.97676	.97681	.97687	97692	.97698	.97703	97709	97714	97720	1 99
	0	1	2	3	4	5	6	7	8	9	

ANTINORMITS

y	Thousandths 0	1	2	3	4	5	6	7	8	9	*y*
2.00	.97725	.97730	.97736	.97741	.97746	97752	.97757	97763	.97768	.97773	2.00
2.01	97778	.97784	97789	.97794	97800	97805	97810	.97815	97820	97826	2.01
2.02	97831	.97836	97841	97846	97851	.97857	.97862	97867	97872	97877	2.02
2.03	97882	97887	.97892	.97897	97902	.97907	97912	97917	97922	97927	2.03
2.04	97932	.97937	.97942	.97947	.97952	97957	97962	97967	97972	97977	2.04
2.05	97982	.97987	97992	97996	98001	98006	98011	.98016	98020	98025	2.05
2.06	98031	98035	.98040	98044	98049	98054	98059	.98063	.98068	98073	2.06
2.07	98077	98082	.98087	98091	.98096	.98101	.98105	.98110	.98115	.98119	2.07
2.08	98124	.98128	.98133	.98137	98142	98147	98151	98156	98160	.98165	2.08
2.09	98169	.98174	98178	98183	98187	98191	98196	98200	.98205	98209	2.09
2.10	98214	98218	98222	98227	98231	98235	.98240	98244	98248	98253	2.10
2.11	.98257	.98261	98265	98270	98274	.98279	98283	.98287	98291	98295	2.11
2.12	98300	98304	98308	98312	98316	.98321	.98325	98329	98333	98337	2.12
2.13	.98341	98346	.98350	98354	98358	98362	98366	98370	98374	98378	2.13
2.14	98382	.98386	98390	98394	98398	98402	98406	98410	98414	98418	2.14
2.15	.98422	98426	98430	98434	98438	98442	98446	98450	.98454	.98457	2.15
2.16	98461	98465	.98469	98473	98477	98481	.98484	98488	98492	98496	2.16
2.17	98500	98503	98507	98511	98515	98518	.98522	98526	98530	.98533	2.17
2.18	98537	98541	.98545	98548	98552	98556	.98559	98563	98567	98570	2.18
2.19	98574	98577	.98581	98585	98588	98592	.98595	98599	98603	98606	2.19
2.20	.98610	.98613	98617	.98620	98624	98627	98631	98634	98638	98641	2.20
2.21	98645	.98648	98652	98655	98659	98662	98665	98669	98672	98676	2.21
2.22	.98679	98682	98686	98689	98693	98696	98699	98703	98706	.98709	2.22
2.23	98713	98716	98719	98723	98726	98729	.98732	98736	98739	.98742	2.23
2.24	.98745	98749	.98752	98755	.98758	98762	98765	98768	98771	.98774	2.24
2.25	.98778	.98781	98784	98787	.98790	98793	98796	98800	98803	98806	2.25
2.26	98809	98812	.98815	98818	98821	98824	98827	98830	98834	98837	2.26
2.27	98840	98843	98846	.98849	98852	98855	98858	98861	98864	98867	2.27
2.28	98870	.98873	.98876	98878	98881	98884	.98887	98890	98893	98896	2.28
2.29	98899	98902	98905	98908	.98910	98913	.98916	98919	98922	98925	2.29
2.30	98928	98930	.98933	98936	98939	98942	98944	98947	98950	98953	2.30
2.31	.98956	98958	.98961	98964	98967	98969	98972	98975	98978	98980	2.31
2.32	98983	98986	98988	98991	98994	98996	98999	99002	99004	99007	2.32
2.33	99010	99012	.99015	99018	99020	.99023	.99025	.99028	.99031	99033	2.33
2.34	99036	99038	.99041	.99044	99046	99049	99051	99054	.99056	.99059	2.34
2.35	99061	99064	.99066	99069	99071	99074	99076	99079	99081	.99084	2.35
2.36	.99086	.99089	.99091	99094	99096	.99098	.99101	99103	99106	99108	2.36
2.37	99110	.99113	99115	99118	99120	99122	99125	99127	99130	99132	2.37
2.38	99134	.99137	99139	99141	99144	99146	.99148	99151	99153	.99155	2.38
2.39	99158	.99160	.99162	.99164	99167	99169	99171	99174	99176	99178	2.39
2.40	99180	99182	99185	99187	99189	99191	.99194	99196	99198	99200	2.40
2.41	99202	99204	99207	99209	99211	99213	99215	99218	99220	99222	2.41
2.42	99224	99226	99228	99230	99232	99234	99237	99239	99241	99243	2.42
2.43	99245	99247	99249	99251	.99253	99255	99257	.99260	.99262	.99264	2.43
2.44	99266	.99268	99270	99272	.99274	99276	.99278	99280	.99282	99284	2.44
2.45	.99286	99288	99290	99292	99294	99296	99298	99299	99301	99303	2.45
2.46	.99305	.99307	99309	99311	99313	99315	99317	99319	99321	99322	2.46
2.47	.99324	.99326	.99328	99330	99332	99334	.99336	.99338	99339	99341	2.47
2.48	.99343	.99345	.99347	99349	99350	99352	99354	99356	99358	99359	2.48
2.49	.99361	99363	.99365	.99367	.99368	.99370	99372	.99374	.99376	.99377	2.49
	0	1	2	3	4	5	6	7	8	9	

(*Continued from page 536*)

[12] Finney, D J., and Stevens, W. L., "A table for the calculation of working probits and weights in probit analysis," *Biometrika*, 35 (1948), 191–201

[13] Finney, D. J., *Statistical Method in Biological Assay*, London: Charles Griffin and Company, Limited, 1952.

[14] Gaddum, J H., *Reports on Biological Standards, III. Methods of Biological Assay Depending on a Quantal Response*, Medical Research Council, Special Report Series No. 183, London: His Majesty's Stationery Office, 1933.

[15] Gaddum, J. H., "Bioassays and mathematics," *Pharmacological Reviews*, 5 (1953), 87–134. See pp. 114–16.

[16] Gaddum, J H , "Statistical Method in Biological Assay," by D J Finney, *Journal of the Royal Statistical Society*, Series A, 116 (1953), 85–6 See p. 86.

[17] Galton, Francis, "Grades and deviates (including a table of normal deviates corresponding to each millesimal grade in the length of an array, and a figure)," and Sheppard, W. F , "Table of deviates of the normal curve," *Biometrika*, 5 (1906–07), 400–06

[18] Irwin, J. O. and Cheeseman, E. A , "On the maximum-likelihood method of determining dosage-response curves and approximations to the median-effective dose, in cases of a quantal response," *Supplement to the Journal of the Royal Statistical Society*, 6 (1939), 174–85

[19] Kelley, T. L , *Statistical Method*, New York· Macmillan Co , 1923

[20] Kelley, T. L , *The Kelley Statistical Tables* New York: Macmillan Co , 1938.

[21] Lowan, Arnold N., Technical Director, *Tables of Probability Functions*, Vol. II, Federal Works Agency, Work Projects Administration for the City of New York, 1942

[22] Pearson, Karl, Editor, *Tables for Statisticians and Biometricians*, Ed. 2, Part I, London· issued by the Biometric Laboratory, University College, 1924

[23] Pearson, Karl, Editor, *Tables for Statisticians and Biometricians*, Ed. 1, Part II London: issued by the Biometric Laboratory, University College, 1931.

[24] Plackett, R L and Hewlett, P S , "Quantal responses to mixtures of poisons," *Journal of the Royal Statistical Society*, Series B, 14 (1952), 154–63. See discussion following the main paper.

[25] Neyman, Jerzy, "Contribution to the theory of the χ^2 test," *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, (edited by Jerzy Neyman). Berkeley and Los Angeles, University of California Press, 1949, pp. 239–73.

[26] Rao, C. R., "Information and the accuracy attainable in the estimation of statistical parameters," *Bulletin of the Calcutta Mathematical Society*, 37 (1945), 81–91.

[27] Taylor, W. F , "Distance functions and regular best asymptotically normal estimates," *Annals of Mathematical Statistics*, 24 (1953), 85–92.

[28] Urban, F M., "Die psychophysischen Massmethoden als Grundlagen empirischer Messungen (continued)," *Archiv für die gesamte Psychologie*, 16 (1910), 168–227.

NUMERICAL ANALYSIS RESEARCH UNPUBLISHED STATISTICAL TABLES

D. TEICHROEW*
University of California, Los Angeles

THIS note lists tables connected with probability and statistics which were computed at what is now Numerical Analysis Research, University of California, Los Angeles and before July 1, 1954 was the Institute for Numerical Analysis, National Bureau of Standards

Publication of tables is expensive and funds for the publication of statistical tables are extremely hard to find It is therefore likely that many of the tables listed below will not be published Many of them are in fact not yet ready for publication, and some are too special ever to warrant publication It seems worthwhile to report the existence of these tables and indicate their present state

Some of the tables were computed as ends in themselves, others were by-products in computations or were computed because the machine codes were available from other problems In general, the tables have not been verified beyond the accuracy required for the purpose for which each was intended It seems worthwhile, however, to list tables which are not completely verified because they may simplify the checking procedures if the functions are recomputed.

The number of digits given in some of the tables may seem excessive to many statisticians The number of digits used is a consequence of the fact that most of the tables were computed on the SWAC (Bureau of Standards Western Automatic Computer). This machine operates on a basic number length of 36 binary digits (10.8 decimal digits) and there is no point in using less than an integral multiple of the basic length and, in general, the tables retain most of the digits used in the computation.

Publication appears probable for the following three tables and they are therefore not listed elsewhere in this note

1. "Tables of the bivariate normal distribution function and related functions." Collated by the National Bureau of Standards. The introduction is by G. Blanch.

* The preparation of this table was sponsored (in part) by the Office of Naval Research, USN and the School of Aviation Medicine, USAF The author's present address is National Cash Register Company, Hawthorne, Calif.

The functions given are

$$L(h, k, r) = \frac{1}{2\pi\sqrt{1-r^2}} \int_h^\infty dx \int_k^\infty \exp\left[-\frac{1}{2}\left(\frac{x^2+y^2-2rxy}{1-r^2}\right)\right] dy$$

and

$$V(h, \lambda h) = \frac{1}{2\pi} \int_0^h dx \int_0^{\lambda x} \exp\left[-\frac{1}{2}(x^2+y^2)\right] dy$$

2. "Tables of salvo kill probabilities for square targets" *Applied Mathematics Series 44, National Bureau of Standards* Introduction by A D Hestenes.[1]

This table gives values of the function

$$P = \frac{1}{2\pi\sigma_{A_x}\sigma_{A_y}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} Q(\zeta, \eta) \exp\left[-\frac{1}{2}\frac{(\zeta - x_0)^2}{\sigma^2_{A_x}} - \frac{1}{2}\frac{(\eta - y_0)^2}{\sigma^2_{A_y}}\right] d\zeta d\eta$$

where

$$Q(\zeta, \eta) = 1 - [1 - P_k P_R(\zeta, \eta)]^N; \qquad 0 \leqq P_k \leqq 1$$

and

$$P(\zeta, \eta) = \frac{1}{2\pi\sigma_{R_x}\sigma_{R_y}} \int_{-a}^{a} \int_{-a}^{a} \exp\left[-\frac{1}{2}\frac{(x-\zeta)^2}{\sigma^2_{R_x}} - \frac{1}{2}\frac{(y-\eta)^2}{\sigma^2_{R_y}}\right] dxdy.$$

3 "Empirical power functions for nonparametric two-sample tests for small samples." D Teichroew. Accepted for publication in the *Annals of Mathematical Statistics*

This paper gives the empirical frequencies of all possible rankings which are obtained when a sample of m from a normal population of zero mean and unit variance and a sample of n from a similar population but of mean-δ are ranked in order of size, for $(m, n) = (3, 2)$ $(3, 3)$ $(4, 2)$ and $(4, 3)$ and various values of δ.

All the tables listed below exist on punched cards In addition, two tables mentioned in section II, namely, those containing $y(p, \lambda)$ and $Q_s(p)$, were multilithed and a number of copies have been distributed

[1] This table has been published since this note was written.

It is hoped that these two tables will eventually be published in one book.

The tables are listed under five categories:

I. Tables associated with the normal distribution.
II. Tables associated with the Gamma distribution.
III. Tables associated with the t-distribution
IV. Tables for selecting samples from certain distributions.
V. Miscellaneous tables.

I TABLES ASSOCIATED WITH THE NORMAL DISTRIBUTION

Let

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad F(x) = \int_{-\infty}^{x} f(t)dt,$$

$$H(x, b) = \int_{-\infty}^{x} [F(t)]^b dt,$$

$$\psi(a, b) = \int_{-\infty}^{\infty} [F(x)]^a \int_{x}^{\infty} [1 - F(y)]^b dy dx,$$

$$E(x_j; N) = \frac{N!}{(j-1)!(N-j)!} \int_{-\infty}^{\infty} xf(x)F(x)^{j-1}[1-F(x)]^{N-j}dx,$$

$$E(x_j^2; N) = \frac{N!}{(j-1)!(N-j)!} \int_{-\infty}^{\infty} x^2 f(x)[F(x)]^{j-1}[1-F(x)]^{N-j}dx,$$

$$E(x_i, x_j; N) = \frac{N!}{(i-1)!(j-i-1)!(N-j)!} G(i-1, N-j, j-i-1),$$

$$G(m, n, p) = \int_{-\infty}^{\infty} \int_{-\infty}^{y} xyf(x)f(y)[F(x)]^m [1-F(y)]^n [F(y)-F(x)]^p dxdy,$$

and

$$K(\delta; \alpha, \beta) = (\beta + 1) \int_{-\infty}^{\infty} [F(x + \delta)]^\alpha [1 - F(x)]^\beta f(x)dx.$$

The range and size of the tables is as follows:

Function	Range	Decimal Places Tabulated	Accuracy
$f(x)$	$x = -12.00(.02)12\ 00$	32	$\geqq 27$
$[F(x)]^k$	$x = -12\ 00(.02)12.00$ $k = 1(1)19$	32	$\geqq 25$
$H(x, b)$	$x = -12\ 00(.02)12.00$ $b = 1(1)19$	32	$\geqq 25$
$\psi(a, b)$	$a, b = 1(1)19$	32	$\geqq 25$
$E(x_j, N)$	$j = 1(1)N$ $N = 1(1)22$	21	$\geqq 19$
$E(x_j^2, N)$	$j = 1(1)N$ $N = 1(1)21$	21	$\geqq 19$
$E(x_i, x_j; N)$	$i, j = 1(1)N$ $N = 1(1)20$	21	$\geqq 19$
$K(\delta; \alpha, \beta)$	$\delta = -3\ 2(.1)0(\ 01)6\ 4$ $\alpha = 1(1)9$ $\beta = 0(1)4$	8	8

II. TABLES ASSOCIATED WITH THE GAMMA DISTRIBUTION

Two SWAC routines have been developed for computing $p(y, \lambda)$ and $y(p; \lambda)$, where

$$p(y, \lambda) = \frac{1}{\Gamma(\lambda)} \int_0^{y(p,\lambda)} e^{-t} t^{\lambda-1} dt.$$

1. The first computes p when y is given for λ an integer by summing the series

$$p(y, \lambda) = 1 - e^{-y}\left(1 + y + \frac{y^2}{2!} + \frac{y^3}{3!} + \cdots \frac{y^{\lambda-1}}{(\lambda - 1)!}\right)$$

The routine is also used to compute y for a given p by inverse interpolation

2. The second routine computes y for a given p by the asymptotic series (using terms up to and including Q_{11})

$$y(p;\lambda) = c^2 + Q_1(p)c + Q_2(p) + \frac{Q_3(p)}{c} + \cdots$$

where $\lambda = c^2$. The form of the Q functions was given by Campbell (1923).
These routines have been used to compute the following tables:

Function	Range	Decimal Places Tabulated	Accuracy
$p(y;\lambda)$	$y=0(\ 5)$ varying, $<64\ 0$ $\lambda=2(1)\ 20$	10	5–10
$y(p;\lambda)$	$p=\ 000(\ 001)\ 999$ $\lambda=2(1)15,\ 20(10)50,\ 100$	8	4–8
$2y(p,\lambda)$	$p=.000(\ 001)\ 999$ $2\lambda=3(1)5(2)29$	7	4–7
$Q_i(p)$	$p=\ 500(\ 001)\ 999$ $i=1(1)11$	8	8

III TABLES ASSOCIATED WITH THE t DISTRIBUTION

The tables are concerned with solving the equation

$$p = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\,\Gamma\left(\frac{n}{2}\right)} \int_{-\infty}^{t(p,n)} \frac{d\tau}{\left(1 + \frac{\tau^2}{n}\right)^{(n+1)/2}}$$

for $t(p, n)$ when p and n are given An approximation to $t(p, n)$ is obtained by summing the first 8 terms of the asymptotic series

$$t(p,n) = x + \frac{H_1(p)}{n} + \frac{H_2(p)}{n^2} + \cdots.$$

The first four H functions were determined by Hotelling and Frankel [2]. Their method was used to get H_5, H_6, H_7 and H_8

The tables available are the following:

Function	Range	Decimal Places Tabulated	Accuracy
$H_i(p)$	$p = .500(.001).999$ $i = 1(1)8$	8	8
$t(p; n)$	$p = .500(.001)999$ $n = 2(1)\ 16, 20, 25, 50, 100, 200$	8	2–8

IV TABLES FOR SELECTING SAMPLES FROM CERTAIN DISTRIBUTIONS

One method of selecting samples from distributions consists of the following stesps·

1. Compute, $\theta(k)$, the sum of k variates uniformly distributed on (0, 1)
2. Let

$$x = \sum_{i=0}^{m} a_i \theta^i.$$

The a_i are chosen so that x has the required distribution (This method is developed in Teichroew (1953)). These coefficients have been computed for the case where $k=8$ and for the following distributions:

1. The normal distribution
2. $t/\sqrt{n}$, where t has a normal distribution, for $n=50(1)200$
3. The Gamma distribution, for $\lambda=2(1)\ 15, 20, 25, 50, 100.$
4. $1/\sqrt{y}$ where y has a Gamma distribution for $\lambda=2(1)\ 15, 20, 25, 50, 100$

The variates $t/\sqrt{n}$ and $1/\sqrt{y}$ have been used to generate random values of the inverses of Wishart matrices (If y has a Gamma distribution with parameter λ, then $2y$ has a Chi-Square distribution with 2λ degrees of freedom)

V. MISCELLANEOUS TABLES

1 Multinomial coefficients.

This table gives the function

$$\frac{(m+n+p+2)!}{m!n!p!}\cdot$$

for all combinations of m, n, and p such that $m+n+p \leq 18$ The func-

tion occurs in the expression for the covariances of order statistics.

2. Coefficients for curve fittings by Chebyshev polynomials.

This table gives 10 decimal digit values of

$$\cos \frac{k\pi}{2n} (2\alpha + 1)$$

for $n=2(2)$ 12, 16, 18, 20, 25, 30, 40; $k=1(1)$ $n-1$, $\alpha=0(1)$ $n-1$ This function appears in the curve fitting method described, for example, in *Tables of Chebyshev Polynomials $S_n(x)$ and $C_n(x)$*, National Bureau of Standards, Applied Mathematics Series 9 Introduction by Cornelius Lanczos.

3. Tables for Probit Analysis with Poisson Error Models.

Three functions have been tabulated:

$$U(\alpha, \beta, h) = \sum_{d=1}^{\infty} \frac{e^{-h}h^d}{d!} \int_{-\infty}^{\alpha+\beta \log_{10} d} \frac{e^{-t/2}}{\sqrt{2\pi}} dt$$

$$Z(\alpha, \beta; h) = \sum_{d=1}^{\infty} \frac{e^{-h}h^d}{d!} \frac{e^{-1/2(\alpha+\beta \log_{10} d)^2}}{\sqrt{2\pi}}$$

$$T(\alpha; \beta; h) = \sum_{d=1}^{\infty} \frac{e^{-h}h^d}{d!} (\log_{10} d) \frac{e^{-1/2(\alpha+\beta \log_{10} d)^2}}{\sqrt{2\pi}}$$

for $h=1(1)$ 17, $\beta=0(.5)$ 10, $\alpha=-5(.5)$ 5. The accuracy varies from 3 to 8 decimal places

REFERENCES

[1] Campell, G. A., "Probability curves showing Poisson's exponential summation," *Bell System Technical Journal*, 2 (1923), 95–113.

[2] Hotelling, Harold, and Frankel, L R. (1938), "The transformation of statistics to simplify their distribution," *Annals of Mathematical Statistics*, 9 (1938), 87–96.

[3] Teichroew, D., "Distribution Sampling with high speed computers," Ph D Thesis, University of North Carolina, 1953.

A REFINEMENT IN THE USE OF MARK-SENSE CARDS FOR TEST RESEARCH

VALENTINE APPEL
Nowland and Company, Greenwich, Conn.
AND
GEORGE COOPER
Statistical Tabulating Company, New York, N. Y.

A MAJOR deterrent to the wider use of punched card methods in the analysis of psychological test data is the time-consuming and costly procedure which is necessary in order to punch the item responses into the cards For this reason, most researchers have preferred to use the IBM test scoring machine for purposes of test scoring and item analysis.

The use of the test scoring machine is limited, however, whenever the statistical design becomes complex, or the sample size becomes especially large. This limitation derives from the necessity to hand sort all test papers, and subsequently to feed them one by one into the scoring machine. The speed at which analyses can be accomplished is, therefore, limited by the rapidity with which an operator can manually shuffle the test papers and then feed them into the machine.

Punched card procedures do not suffer from this limitation, and recently techniques and devices have been developed which obviate the necessity for the time consuming key punching procedure. One of these devices is the Document-to-Card Punch which has been developed by the Personnel Research Branch of the Adjutant General's Office. "This device consists of several components—first, a test scoring machine chassis containing a sensing unit and plug board. The IBM answer sheet is fed into the hopper of this machine. Then a tabulating card, with punched holes corresponding to the item responses, is prepared by the second component, which resembles a modified IBM reproducer. Unique identifying information for each answer sheet is transcribed into the punched card concurrently by means of the third component, a manually operated keyboard" [3, p. 155] Unfortunately, this machine is not generally available for non-government work.

Several attempts have been made to adapt mark-sense cards as examination answer sheets in order to permit the examinees to record their responses directly on to a punch card. One of these applications was described in an article by Gage and Remmers [2] in which mark-sense cards were used in a self-administered student opinion poll, and

another was mentioned in a recent article by Staugas [5] in which he described a method for scoring tests from punched cards. The present article describes another application of the use of mark-sense cards in test research. This application outlines a procedure which is not limited by the 27 column capacity of the standard mark-sense card, and which incorporates a machine checking procedure which permits the editing of multiple-coded columns

The study to be reported here involved the administration of a 432 item True-False questionnaire to a group of 500 college students Because considerable item intercorrelation was anticipated, it was regarded as preferable to have the responses punched into IBM cards for later item analysis. In order to save the time and cost of manually key punching and verifying these questionnaires, and because the group to be tested was an intelligent one, it was decided to use mark-sense cards and to allow the examinees to record their responses directly.

Three cards were used for each examinee as follows. One box of standard IBM mark-sense stock was overprinted as shown in Figure 1 [1] Three decks of 500 overprinted cards were then separately punched with consecutive numbers (001–500) using a sequence numbered deck in the reproducing punch At the same time each deck was gang-punched with the letter *S*, *M*, or *V*, which corresponded to a similarly labeled section of 144 items in the questionnaire The three decks were interpreted and then merged on the collator in sequence 001 *S*, 001 *M*, 001 *V*, 002 *S*, 002 *M*, 002 *V*, etc , resulting in 500 sets of three cards each At the time of the administration of the questionnaires each examinee was supplied with an electrographic pencil and one set of three mark-sense cards, on the backs of which the examinees recorded their names and other pertinent background data.

Because of possible inconsistencies in the examinees' marking of the cards, and because of possible machine errors in converting the electrographic pencil marks into punches, it was essential that a procedure be established whereby such discrepancies could be isolated and corrected Ordinarily this can be done automatically by means of the IBM mark-sense reproducer's multiple punch and blank column detector, which rejects all cards having more or less than one punch in a column But because of the fact that six items were coded into each column, this procedure was not possible. For this reason, after all the questionnaires were administered, the three decks of cards containing

[1] The writers wish to express their appreciation to Mr Herman Greenblatt, formerly of Richardson, Bellows, Henry & Co , who was responsible for the card layout and the overprinting

the mark-sensed item responses were passed several times through the IBM reproducer in order to convert the electrographic pencil marks into punches The IBM Electronic Statistical Machine, Type 101, was then wired so as to reject any card which did not conform to the following pattern of six punches in each of the 24 marked columns· X or Y, 0 or 1, 2 or 3, 4 or 5, 6 or 7, and 8 or 9. As each card was rejected it was edited and the procedure was iterated until all of the cards conformed to the prescribed pattern Although actual records were not kept at the time, roughly between 10 per cent and 20 per cent of the

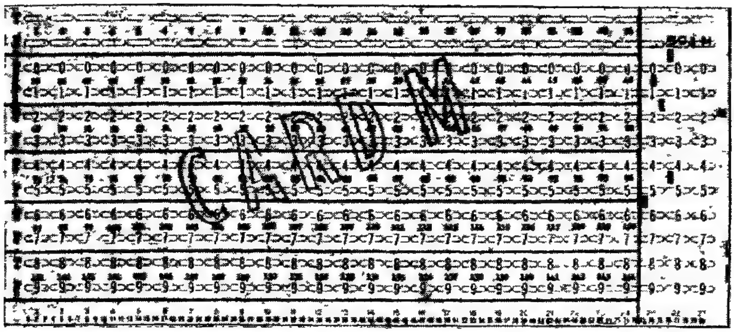


FIG 1 Example of a Mark-Sense IBM Card Overprinted as an Answer Sheet.

1500 cards required some editing in order to make them conform to the prescribed pattern. Most of these editorial changes resulted from occasional item omissions which were arbitrarily coded by the editor as "false." Of the total number of examinees to whom the questionnaire was administered seven cases had to be eliminated from the sample because of failure to follow instructions.

Once all the cards were properly edited, they were sorted into the examinee's code number sequence, each of the three decks separately The first deck was reproduced into regular IBM card stock The second and third decks were then successively reproduced into this new card, each time comparing on the examinee's code number to make certain that each examinee's three mark-sense cards had been properly integrated into the one new card In this way, one punched card was created for each examinee, each card having 432 True-False questionnaire items punched into 72 columns, six items to a column The remaining eight columns were used for the examinee's identification code number and other pertinent information This deck of master

cards was then employed for the purpose of item analysis, and for later scoring and statistical analysis.

Although the study reported here employed a card format which only permitted six two-alternative items to each card column, the method is by no means limited to this. Mark-sense cards may easily be overprinted in any way desirable, and any or all of the twelve marking positions in each column may be employed. While this procedure is not applicable to all test research problems, there are many situations in which the use of a mark-sense card in place of an answer sheet can be most economical of time and effort. Situations in which the method may be used to its greatest advantage are those in which the group to be tested is fairly cooperative and intelligent, in which the sample size is large, and in which complex item analyses are to be performed.

REFERENCES

[1] Deemer, W. L. Jr., "The use of mark sensing in a large scale testing program," *Journal of the American Statistical Association*, 43 (1948), 40–52.

[2] Gage, N. L., and Remmers, H. H., "Opinion polling with mark-sensed punch cards," *Journal of Applied Psychology*, 32 (1948), 88–91.

[3] Harmon, Harry H., and Harper, Bertha P., "AGO machines in test analysis," *Proceedings, 1953 Invitational Conference on Testing Problems*, Princeton: Educational Testing Service, 1953.

[4] Kroeger, Arthur, "A device for simplifying tabulation," *The Journal of Marketing*, 18 (1954), 285–7.

[5] Staugas, Leonard, "A rapid method for scoring tests punched in IBM cards," *Educational and Psychological Measurement*, 14 (1954), 101–5.

SUMMARIES OF PAPERS DELIVERED AT THE 114th ANNUAL MEETING OF THE AMERICAN STATISTICAL ASSOCIATION IN MONTREAL SEPTEMBER 10–13, 1954

Edited by ARMEN A. ALCHIAN, *University of California, Los Angeles*

The present section contains all available abstracts of papers presented at the 1954 national meeting of the American Statistical Association in Montreal, Canada. The sequence of presentation here is alphabetical according to author.

PAPERS SUMMARIZED

On the Accuracy of Short-Term Entrepreneurial Expectations. OSKAR ANDERSON JR., RAINALD K. BAUER, and EBERHARD FELS.

During the past four years, the Ifo Institute for Economic Research in Munich, Germany, has gathered monthly information on anticipated and actual changes of economic microdata as seen by businessmen. The data are not numerical but indicate trends only.

In this study, the data of 104 producers of outer garments, 110 textile manufacturers, 77 textile wholesalers, and 442 textile retailers have been evaluated with a view to the accuracy of their predictions as compared with their ensuing reports on actual changes. The results of this micro-approach are later compared with the findings as they are usually represented in macro-form in the so-called Business Mirrors of the Ifo Institute

First it is found that industrialists appear to have one-month prediction horizons in mind, although the pertinent questions refer to two-month intervals.

Second, the percentage of businessmen's correct estimates increases, once a persistent, uniform trend is under way Otherwise, entrepreneurs have a seeming predilection for giving indifferent expectations For prices, however, they are likely to exaggerate forthcoming upward movements and minimise downward movements in their predictions, in stating actual changes, the reverse holds true.

All findings are subjected to detailed investigations with regard to the various sub-branches and the sizes of the firms The problem of dealing with homogeneous sub-sets of data is also treated Systematic errors in entrepreneurial reporting habits are investigated and a suitable graphical form is given for illustrating the decisive "indifference intervals."

Finally, it is found that accurate-prediction frequencies are far better when regarded individually than when treated summarily in aggregates.

A Graduate Course in Basic Statistical Analysis for Majors or Minors in Statistics. R L ANDERSON

This paper discusses a two-semester course in basic statistical analysis, taught for the first time at the University of North Carolina last year The course was designed to parallel a similar course in basic statistical theory. Most of those students who took both courses seemed to do very well, however, students taking only the analysis course without a previous course in theory were severely handicapped

The course was divided into three parts descriptive methods, sampling experiments, and analysis of data and design of experiments and surveys The first part took too long and did not interest the students Improved descriptive tools are needed, and better methods of teaching them, at least better than were used in this first attempt.

The sampling experiments were designed to present the basic ideas of statistical inference by drawing samples from known populations For many of the students, the tedium of computing tended to obscure the main objectives of the study More time needs to be spent in preliminary lectures and on complete sampling from small populations to demonstrate the meaning of expectations. We found that students were able to utilise the results of large scale sampling outside of class, e g , IBM sampling After some revision of the teaching procedure, I feel that the use of empirical sampling offers real promise in presenting the ideas of statistical inference It is important to emphasise the need for this when theory is not available

Even though the analysis and design part of the course was curtailed, the results seemed to be highly satisfactory. It would be desirable to have some impartial observer conduct an examination on analysis and design to find out if the students learned general principles or only those procedures mentioned by the teacher

In order to have time to present the important methods of collecting and analysing data, the following changes are suggested

(1) Cut down the time devoted to the introduction by distributing mimeographed materials and requiring outside reading.
(2) Condense the descriptive materials.
(3) Have results of large scale sampling experiments available before the class work begins
(4) Introduce empirical sampling by drawing all possible samples from small populations
(5) Be sure lectures precede the sampling, so that students have a preview of the purposes of empirical sampling.
(6) Schedule 3 lecture hours and one three-hour supervised laboratory each week Emphasize that some non-supervised laboratory work will also be expected.

One final comment If an analysis course is designed to parallel a similar course in theory, all students should either be taking the theory course or have had it already Otherwise, it is imperative that a certain amount of theory be taught in the analysis course.

The Criticism of Transformations. F. J ANSCOMBE and JOHN W. TUKEY, *University of Cambridge and Princeton University.*

The classical methods of analysis of variance and regression are flexible, relatively easy to apply, and far more widely familiar than any others. When they do not apply directly and it is possible to apply them wisely to transformed responses, it is almost always better to bend the data to the analysis than to bend the analysis to the data Purposes of transforming data are (1) to increase the additivity or simplicity of treatment effects (generally the most important purpose, affecting the quality of inferences to be made from the data), (2) to make the variance more constant, and (3) to reduce non-normality Two general types of criticism of transformations on the basis of experimental data are (*a*) the running check, experiment by experiment, to detect cases of extreme non-additivity, non-constancy of variance and non-normality, and (*b*) the serious study of a particular field of experimentation to determine what transformations are most reasonable for routine use in that field For (*b*), the data from a number of experiments will need to be analysed according to many (perhaps 5, 10 or 20) transformations. Families of transformations form spaces with structure determined by their statistical properties The conclusion of study (*b*) will be to specify a confidence region in the space of transformations for the satisfactory transformation

The procedure suggested for testing the adequacy of any transformation is essentially as follows (but many variations of detail are possible) Corresponding to each transformed observation y calculate the fitted value Y given by least squares and the residual $z = y - Y$. Plot the z's against the Y's Removable non-additivity appears as a curved regression, removable non-constancy of variance appears as a wedge-shaped outline, and extreme non-normality is reflected in non-normality of the marginal distribution of the z's If it is desired to supplement this graphical analysis by calculated tests or estimates, the following statistics may be used

$$\text{(i) } \sum zY^2, \qquad \text{(ii) } \sum z^2Y, \qquad \text{(iii) } \sum z^3, \qquad \text{(iv) } \sum z^4,$$

where (i) is for non-additivity (see Tukey, *Biometrics*, 5 (1949), 232-42), (ii) is for non-constant vraiance, and (iii) and (iv) are for skewness and kurtosis

Statistics and Planning Educational Operations C M ARMSTRONG, *New York State Education Dept*

The statistics that should be available for administration in educational institutions are inadequate In some cases the inadequacy is the result of lack of any records at all and in other cases it is the result of poorly designed or poorly executed reporting plans

Some of the areas in which statistics are most essential are (1) determining the size and location of school buildings, (2) planning the organization of the teaching force, (3) determining the teacher demand and supply, (4) planning for the guidance of students, (5) planning the curriculum and (6) establishing and maintaining standards.

Areas particularly needing improvement are, (1), (3), (4), and (6)

An important improvement is to reconcile and improve school census data and United States Census data so that they can be used inter-changeably The present inconsistencies are confusing to those trying to use the data

In the areas of child guidance and establishing and maintaining standards, a great deal of research work is needed to devise techniques for processing data on child characteristics and growth so that the actual changes in the school children can be used more effectively as guides to administrative action With more attention to these statistics there is a strong possibility that a system of controls for school processes might be established that would correspond with the statistical quality control techniques used in manufacturing.

The Development of Census Tracts in the United States. C. E BATSCHELET, *Census Bureau.*

The first census tracts were established in eight of the larger cities of the United States in connection with the Census of 1910 Since that time the census tract program has developed at an accelerated rate until now there are tracts in 142 of the 238 cities of 50,000 or more population In the tracted areas there are 14,500 tracts, accounting for a population of 62 million, or 41 1% of the total population of the United States.

Census tracts are established by interested local groups on the basis of principles defined by the Bureau of the Census. In the 1910, 1920, and 1930 Censuses, the tracted cities had to meet the expense of the tract tabulations, in the 1940 and 1950 censuses the Bureau of the Census compiled and published population and housing data by census tracts as part of the regular reports

It was in 1930 that Howard Whipple Green of Cleveland first convinced the Bureau of the Census that the areas adjacent to the central tracted cities also should be tracted By the 1940 Census, 25 cities had established tracts in their adjacent suburban areas and the Bureau of the Census decided that it would be desirable to extend the tract program to include the metropolitan districts of all of the tracted cities In the 1950 Census, standard metropolitan areas, consisting of whole counties, were adopted for

the presentation of statistics in lieu of metropolitan districts, and the tracting of the standard metropolitan areas was adopted by the Bureau of the Census as a desirable goal. At present, there are 172 standard metropolitan areas of which 47 are tracted in their entirety and 42 have the central city or the city and part of the adjacent area tracted.

The Bureau of the Census has just completed an annotated bibliography, *Census Tract Publications Since 1950*, compiled from information received from key persons in the tracted areas The bibliography demonstrates the extensive and diverse uses which are made of census tracts in marketing analysis, demographic, public health, and social programs, and planning and research activities It is expected that the bibliography will further stimulate the use of census tracts in the localities in which tracts have been established and will show other cities the advantages to be derived from participation in the Census Tract program.

Estimation by Least Squares and by Maximum Likelihood JOSEPH BERKSON, *Mayo Clinic*

The situation dealt with is given by $P_i = F(x_i, \alpha, \beta) = F(Y_i)$ (1), where P_i is the probability of an event corresponding to x_i, α and β are parameters to be estimated, $Y_i = \alpha + \beta x_i$ (2) is the linear transform of (1), $\hat{p}_i = F(x_i, \hat{\alpha}, \hat{\beta})$ (3), and $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$ (4), are estimates of (1) and (2) respectively

A class of least squares estimates is defined by minimization of (a) $\sum w_i (p_i - \hat{p}_i)^2$ or (b) $\sum W_i (y_i - \hat{y}_i)^2$, where $p_i = 1 - q_i$ is an observed relative frequency corresponding to P_i based on n_i observations at x_i, y_i is the linear transform value corresponding to p_i, $1/w_i$ is any asymptotically efficient estimate of the variance of p_i, $1/W_i$ is any asymptotically efficient estimate of the variance of y_i Estimates falling in this class are asymptotically efficient and therefore asymptotically equivalent to the maximum likelihood estimate If F is the logistic function, y_i the logit of p_i and $W_i = n_i p_i q_i$, the estimate obtained is the "minimum logit X^2 estimate", if F is the integrated normal curve, y_i the normit (normal deviate) of p_i, and $W_i = n_i z_i^2 / p_i q_i$ where $z_i = dp_i / dy_i$, the estimate is the "minimum normit X^2 estimate"

The minimum logit X^2 estimate and the minimum normit X^2 estimate were compared with the maximum likelihood estimate for finite samples, in respect of the variance and mean-square-error of the estimates For all situations investigated it was found that these least squares estimates have smaller variance and smaller mean square error than the maximum likelihood estimate

Measuring the Effect of Unemployment Benefits on the Economy. MARVIN K BLOOM, *Research Council for Economic Security, Chicago, Illinois*

The unemployment compensation system in the United States was designed to encourage stabilization of employment and to tide workers over limited periods of unemployment Although there has been some dispute on its role in "maintaining purchasing power", there is increasing interest in the system as an "automatic stabilizer" The system clearly meets three criteria for an "automatic stabilizer" for the following reasons it goes into action automatically, on an annual or quarterly basis since 1940, unemployment compensation (benefits less taxes) added to purchasing power during periods of slump and subtracted from it during high prosperity In the first quarter of 1954, state benefits exceeded state taxes by more than one-third of a billion dollars

It is harder to determine the extent to which unemployment benefits reduce the public's demand for cash in a slump Annual benefits have never exceeded 2 3 per cent of taxable wages or 0 9 per cent of disposable personal income More important is the proportion of income loss offset by unemployment benefits Four broad classes of benefit-income loss measures are presented Two classes are based on gross wage loss, two on net wage loss Recent estimates of the proportion of gross wage loss offset by benefits range from 18 to 25 per cent Estimates of the net income loss offset by increases in benefits in the 1948–1950 decline, based on quarterly data, range from 5 to 25 per cent The analysis is extended to selected labor-market areas to show the differential impact of income changes and benefit payments on different segments of the population Here, the relative offset ranged from 3 to 25 per cent The following factors account for differences in the ratios level of benefits provided, the income base used, method of computing additional benefits, use of unadjusted or seasonally-adjusted figures, the relative income loss, the different components of the income loss

The following concepts are discussed What constitutes a wage loss for the economy? Is net wage loss a better base than gross wage loss? What are the effects of divergent movements in the components of income on consumption? What happens to the benefits that unemployed workers receive? The fragmentary data in this area are summarized Attention is drawn to pilot studies by the U S Department of Labor of the income and expenditures of claimants' families A research program is suggested which coordinates such studies with continuous work histories of samples of covered workers and beneficiaries.

Canada and the Outside World C D. BLYTH, *Dominion Bureau of Statistics, Ottawa*

Inter-relationships between Canada and the United States as shown in the Canadian balance of payments are numerous and far reaching An outstanding difference between Canada and the United States is the relatively greater importance of foreign trade in the case of Canada Foreign transactions

make up about one-quarter of Canada's gross national expenditure compared with around 5 per cent in the case of the United States. And the very large and growing ratio of this Canadian trade is with the United States, Canada being the best customer of that country as well as the leading source of supply. The close connection with United States demand is one of the most direct contacts which the Canadian economy has with the United States. Many of the basic Canadian primary industries supply the United States.

There is also an influential chain of inter-relationships arising from the large number of Canadian industries which are controlled by United States corporations. These constitute about one-fifth of the investment in Canadian industry and commerce and in manufacturing they represent not far from one-third of the capital invested, and make for increasing interpenetrations between the United States and Canadian business communities Security markets are also closely inter-related, and there is also an unprecedented movement of people across the border

But the Canadian economy has a distinctly separate existence to that of the United States. Among the factors differentiating Canada from the rest of North America is affiliation with the Commonwealth and the relatively greater extent of Canadian relations with the overseas world The rapid rate of growth in Canada's domestic economy has also provided increased strength and a momentum to Canadian activity, but external transactions remain of special importance.

In relation to the overseas world, Canada is now one of the "have" countries Canadian participation in the recent war and subsequent loans and contributions to overseas countries in the post-war years illustrate Canada's economic strength, and post-war growth has further added greatly to industrial development. This new situation contrasts with the position of economic dependency in the early decades of the century. While the net foreign indebtedness of Canada has risen since 1950, it is less in real terms than before the war and capital from Canadian sources has financed much the largest part of new investment. But the existence of deficits in Canada's current account in recent years arising from large expenditures in the United States and the continued debtor position of the country have placed limitations on the extent to which Canada can provide overseas aid.

Finally, comment is made on the partly unexplored frontiers of international financial statistics—short-term movements of capital The elusive and complex character of this field has provided a baffling problem of measurement for all countries which are concerned with it

State Patterns and Speed of the Labor-Force Shift from Agricultural to Nonagricultural Industries in the United States, 1870–1950. C. P. BRAINERD, *University of Pennsylvania*

The pattern and rate of the labor-force shift produced by divergent developments in agriculture and in industries are analyzed in terms of the "nonag percentage," or the number of nonagricultural workers in each state taken as a percentage of the total labor force in that state, and used as an index of economic maturity. The dominant course of development is a strong upward movement with no reversals, but three subordinate regional patterns are distinguished within the main one

Nonagricultural workers increased their share of the labor force from 48 to 88 per cent in 80 years. Using point change (increment) to measure rate of change, southern states changed the fastest, central and western states came next, and northeastern states changed most slowly But measuring rate of change in terms of "rate of approach" to a hypothetical maximum of 100 per cent nonag, the industrial states of the northeast and central regions moved fastest towards the limit, while southern and mid-western farm states moved the slowest Although the country as a whole made the transition at an even pace by either measure, a majority of the states changed faster before 1910 than later, but those that changed faster after 1910 made striking gains. Basing the rate of approach on the potential change as of 1910, the transition was much accelerated in the second period The higher a state's level of maturity in 1870, the smaller the net change in the percentage by 1950, but the higher the rate of approach to the maximum Level of start was not, however, a limiting factor until after 1910.

The shift away from agriculture was not directly connected with population growth, but until 1910 was definitely associated with immigration and with rate of growth in number of manufacturing wage earners Decline of the agricultural labor force in the east appears first to have retarded the transition in the west and later to have maintained it in the east In the south, transition was slow until the number of agricultural workers began to decline

Census Data for Units of School Government. HENRY M BRICKELL, *Manhasset Public Schools, New York.*

The basic unit of school government in the United States is the local school district Because of the limitations the typical state has placed upon its own educational authority, educational control and the main burden of educational financing fall within the province of the local district. Accordingly, the person engaged in educational research must look within each school district for the factors which make its schools differ from the schools of other districts.

A dozen of the questions usually asked by a census enumerator give the key to most of the differ-

ences. Adequate follow-up of these leads is being slowed somewhat by the inadequacy of available census data, but slowed much more severely by the fact that information is generally *not available by school districts.*

There are three major factors in the school setting which have arrested our attention. The first is the ability of the people to pay school taxes. We need to know the total value of property and income. The second major factor is population characteristics Past study points to nativity, education, occupation, and other items of personal background as related to the kind of schools operated in the district The third factor demanding study is the type and intensity of group organization within a community

It is not the research man alone who is interested in the census Those responsible for the day to day operation of the schools are year-round users of the kind of data gathered by the censuses of population and housing Moreover, if census data were available by the governmental units with which school officials are concerned, its utility would be multiplied.

Recommendations (1) Data collected by the Population and Housing Division should be published or made readily available by the local governmental units operating public elementary schools (2) The same summary information should be published for all districts regardless of size or location so that comparisons between districts of any size over the nation are possible. (3) In intercensal years local school districts or other governmental units should have access to the standard questions, instructions to enumerators, and canvassing methods used by the Bureau so that they can gather local data comparable to that collected during the Bureau's sample censuses and sample surveys (4) The Bureau should encourage interested organizations to advertise and popularize available data

Statistical Methods in Meteorology GLENN W. BRIER, *U. S. Weather Bureau*

Several examples are given of techniques devised to overcome some of the statistical problems encountered with weather data or other time series. The first topic discussed is the length of record necessary to determine a "normal", and a simple criterion for making a rational choice is given Two other examples are concerned with the problem of testing the significance of "singularities" in weather A description is given of a sampling experiment with random numbers which formed the basis of a test on rainfall data A non-parametric test is proposed for use in testing the significance of the relationship between two time series

The Flow of Funds Approach to Savings and Investment DANIEL H BRILL

This paper is concerned with data requirements for analysis of saving and investment. First, the paper deals with inadequacies in the definitions underlying concepts of saving and investment in the national income accounts. These inadequacies are (a) differences in scope of activities classified as saving from one sector to another, making meaningful comparisons within or among economies very difficult, and (b) consolidation of all investment activities into one account, thereby preventing intersector comparisons and precluding analysis of the influence of financial variables, such as debts or liquid assets, in saving and investment processes.

Next, the paper discusses a new system of accounts, now in preparation, which may be of use in these problems The system, known as "flow of funds accounts", leans heavily on the income and product structure for data The flow of funds system attempts to record all uses of money and credit by each of the major sectors of the economy, whether for goods or services, capital or current account, financial or non-financial activities Each sector's sources and uses of funds are classified in categories of activity carried through the system consistently. No concept of savings or investment is identified in the accounts, detail on different types of transaction is given to permit combinations of data into various formulations as may be required for analysis

Since the system measures flows of funds, internal bookkeeping allocations (such as charges to reserves or interplant transfers within a single enterprise) are not recorded. Further, it is necessary to measure flows of funds at the values at which the flows occur Some internal allocations, particularly those for depreciation or tax reserves, are important determinants of business investment behavior. Differences between book- and market-values may also be important in analyzing such behavior

These deficiencies emphasize the need for development of accounting systems which can encompass internal transactions as well as intertransactor flows, and book- as well as market-values Finally, it is necessary to investigate much more intensively the relationship between proprietors of unincorporated businesses as entrepreneurs and as consumers It is particularly difficult to separate the personal and the business saving of this group, or even to determine whether such separation would be significant analytically

Recent Advances in Government Statistics. ROBERT W. BURGESS, *Bureau of the Census*

Most of the advances to which I shall call attention do not concern subject matter, but consist rather of (1) A better understanding of the way in which good government statistics are and can be used by government, business and social scientists (2) The development of more efficient and economical

methods including, for instance, greater use of administrative records as raw material, greater use of scientific sampling procedures, and employment of electronic computing equipment (3) Better coordination of the statistical results derived by one government agency with related results derived by another agency (4) Prompter collection and publication of results. And (5) More careful definition and computation of familiar statistical measures with more attention to establishing margin of sampling variation and degree of statistical reliability

In addition to these advances of methods and procedures type, the Census Bureau has initiated or is initiating advances concerned with subject matter, as for instance (1) *Water usage by manufacturing concerns* The 1953 Annual Survey of Manufactures is incorporating and the 1954 Census of Manufactures will incorporate material on the amount of water used in a year by manufacturing establishments (2) *Commodity Flow* Managers of sales and production have long wished they had better information on shipments and stocks at various stages of the flow of each commodity from producer to ultimate consumer The Bureau is exploring, for a limited number of commodities, the feasibility of assembling such figures on a useful basis (3) *"Fix-it" Expenditures*—Collection from sample households of monthly expenditures for repairs and alterations, a project intended to fill a conspicuous gap in construction statistics (4) *Statistical relations of establishment, company and enterprise* The Censuses of Manufactures and Business provide data based on establishments, the Bureau of Internal Revenue deals with profits of companies or sometimes, in the case of consolidated returns, of enterprises, sometimes embracing companies in various different lines of activities Determination of aggregate sales, employment and other statistics of establishments, companies and enterprises on a comparable basis, would facilitate an understanding of the structure of our economy and improve present statistics of national income and gross national product I believe that Census Bureau records can provide some valuable material bearing on such interrelations and could provide more by relatively simple extensions of our inquiries on certain points

From Engineering to Engineering Statistics IRVING W BURR, *Purdue University*

Large numbers of engineers and industrial personnel have been taking courses in statistical quality control These courses are of several kinds intensive courses of eight or ten days and part-time courses, either for those in one organization or open to all These may be in colleges or universities, or privately sponsored The trend in training is ever upward, as seen in the growth of the American Society for Quality Control to a membership of over 8,000 in about ten years

There is a place for every possible gradation between those with a single intensive or part-time course, to those with a doctor's degree in mathematical statistics Some of the people most useful to industry are those with a B S in an engineering field, some practical industrial experience and a degree or two in mathematical statistics

Courses for the engineer should stress applications and experiments, particularly the first course and probably also the second Those with greater mathematical maturity will be ready for and welcome considerable theory in subsequent courses

The long-run objective is to get statistical training into all engineering curricula, not only in the form of courses in statistics, but also throughout the engineering and science courses, so that whenever a statistical problem comes up, as in laboratory data or measurement precision, appropriate methods will be used Even when this is accomplished, there will always be need for training those who never went to college or even high school, and also need for additional training for those who want more statistics Such training will, of course, only be a supplement to the person who is studying on "his own" as many do

In the Statistical Class Room IRVING W BURR *Purdue University*

There are three general and interrelated problems in statistics for physical scientists The first is that of selling them on wider use of statistics The fundamental core is that statistics is the science of analyzing problems involving variation Since physical scientists tend to encounter more and more variation as their experiments become more refined, statistics is becoming a basic tool They must also be shown how easily, without proper design in their experimentation, they can reach wrong conclusions and waste much time

The second problem is that of how to teach physical scientists Shall the emphasis be on theory and derivations, upon experiments to illustrate the theory, or upon applications using physical science data? The writer believes that in a first course the emphasis should be upon the latter two, despite the fact that most physical scientists have the mathematical background to "take" theory There is not time for all three phases of teaching Experiments and applications are more likely to foster that precious quality "statistical thinking" This we must have, because the one course may be our only chance

The third problem is that of what to teach Here the aim should be to present significance of differences, interval estimation, analysis of variance, linear correlation, curve-fitting and design of experi-

ments as soon as possible, for these are the tools for which the physical scientist can see a use in his work. The menu can be rounded out in a second course.

Throughout the course the aim should be sell, sell, sell! The more the statistician knows of physical scientists' problems the better he can teach them

Optimal Filtering as Statistical Decision. A GEORGE CARLTON, *The Johns Hopkins University.*

The problem of filtering, i.e, estimating present value of a message random process on the basis of observation of the past of the input process made up of the sum of the message process and a noise process, is considered The Wiener optimal realizable linear filter is exhibited Minimax σ^2 realizable linear filters are considered It is shown that with message and noise spectra given, and process distribution functions only slightly further restricted, the minimax realizable filter is the optimal realizable filter The minimax realizable filter is derived for cases in which either or both of the spectra are subject only to a linear restriction of the form $\Phi(\omega) = C(\omega) f(\omega)$, where $C(\omega)$ is a specified function designated the spectral capability function and $f(\omega)$ is any symmetric probability density function Simple special cases are derived explicitly It is pointed out that in many cases one can design subminimax adaptive filters which possess all the merits of Wiener-type optimal linear filters based on spectra estimated from the observations and furthermore, adapt reasonably well to nonstationary and non-normal processes, and that in some cases these adaptive filters are most naturally constructed in feedback form

Change of Composition Estimates of Fish Populations. DOUGLAS G. CHAPMAN, *Oxford*

The theory of the method of population estimation based on the change of composition of the population due to a selective removal is studied in some detail The author has given the maximum likelihood estimates earlier ("The estimation of biological populations," (*Annals of Mathematical Statistics* 25, 1954), 1–15] under the assumption of random sampling, so that a binomial model is valid Some of the shortcomings that occur in the practical application of this method are here considered in their effect on the estimation procedure In general the procedure appears to be fairly insensitive to failure of the assumptions The method is compared with the more usual tag sample estimation procedure, in terms of the information obtained for the same amount of effort It appears that the tag sample method yields more information but at the same time is more likely to give unsatisfactory results due to failure of the underlying assumptions

Also considered is an estimation procedure based on a combination of tag sample and change of composition methods In one important case at least the maximum likelihood equations are easily solved Some answers are given to questions of optimum sample design where the binomial model is assumed An estimation procedure is outlined for the case where the sampling is such that the binomial model is not acceptable

Statistics in Colleges and Universities of the South H H CHAPMAN

The Inventory of Instruction, Research and Service in Statistics was sponsored by the Southern Regional Education Board Information was collected largely by questionnaires but was supplemented by an examination of college catalogues and by personal visits to institutions A total of 272 institutions in the fourteen states served by the Southern Regional Education Board were invited to participate, questionnaires were sent to 209 and replies were received from 193 A preliminary report gives detailed information on the courses taught in the several institutions, the laboratory and other facilities available, the place that statistics occupies among the elective or required courses in various curricula, programs for training statisticians, research and service activities undertaken, and organization for instruction, research and service in statistics The final report will be published by the Southern Regional Education Board

The data collected indicate a wide diversity of offerings The mathematical preparation for the introductory courses range from no formal requirement to calculus For the more advanced courses a knowledge of calculus is ordinarily required To a very large extent the introductory courses are given in subject matter departments and are designed primarily for students majoring in the given department Centralized statistics departments were reported in only three institutions although some type of centralized control or coordination was reported in several of the other larger institutions The reports indicate that many of the institutions have made substantial progress in providing the library facilities and the equipment needed for advanced graduate instruction and for research and consultation service

An analysis of the programs in operation and the opinions expressed indicates decided differences in thinking concerning the most desirable lines of development Some are primarily interested in the training of statisticians, tend to place the emphasis on statistical theory, and are distressed by the numerous offerings of introductory courses by subject matter departments The majority, however, are most directly concerned with the application of statistical methods to the solution of problems in a chosen subject-matter field.

Pensions—A Stabilizing Influence on Consumption. MIRIAM CIVIC

The present and potential effects of pension payments on the economic stability of the country have received little of the attention due them This paper is an initial step in measuring the actual and relative dimensions of pensions today, and in tracing their growth from a half dozen years ago to a half dozen years hence. It highlights the inadequacy of available statistics for appraising the stabilizing effects of pensions.

Pensions from all sources amounted to $7½ billion in 1953 compared with less than half this sum in 1947. The author estimates that by 1960 they will have doubled Since most pension income is spent, it is interesting to note that pensions represented about 2% of total consumption expenditures in 1947 and a little over 3% in 1953 They will come close to 5% in 1960, under conditions of relatively high employment

Three-fifths of pension outlays went to persons 65 or older The rest were paid to younger persons on account of disability or survivorship. Expanded pension programs since 1947 have had noticeable effects on the sources and amount of income received by the aged, the proportion of older persons "with income", and the number of new households formed by aged individuals Characteristics of the aged as a group, which indicate that pension income will be rapidly spent, include their relatively high net worth and home ownership, favorable tax status, small fixed commitments, and low rate of savings. Empirical data are badly needed to show how the aged actually spend, or save, their income

Inherent in all pension programs is an expansion factor which will be responsive to economic decline It consists of pensionable persons who are able to hold jobs when business conditions are good, but who will leave the labor market during a setback Such a development was observed between January, 1953 and January, 1954, when 200,000 aged workers who were eligible for social security withdrew from the labor force to collect benefits But possible expansion is less important than the certainty that pension income will at least hold steady, when other types of income decline

Some Methods for Strengthening the Common χ^2 Tests. WILLIAM G COCHRAN, *Johns Hopkins University*

Since the χ^2 tests of goodness of fit and of contingency tables are not directed against any particular alternatives to the null hypotheses, it is often advisable to replace or supplement them with more specific tests. This paper gives a review, with illustrative examples, of some common methods of strengthening χ^2 tests in this sense There may be a substantial increase in power in the χ^2 test itself by the use of small expectations at the tails. An example which brings out this point is discussed

An alternative to χ^2 that will often detect departure from the null hypotheses is the comparison of low moments of the observed and theoretical distributions, e g , by means of the "variance" test for the binomial and Poisson distributions and the tests of skewness and kurtosis by the normal distribution The sum of squares in a variance test may also be broken down to give more specialised tests Another possibility is to test some selected linear function of the deviations between observed and expected numbers in a goodness of fit test.

Some recent methods for subdividing the degrees of freedom in the χ^2 test for a two-way contingency table are also relevant. These may be particularly useful when one or both of the classifications in the two-way table is ordered.

Some methods are presented for handling the fairly common problem of combining results from a number of independent two-way tables.

Needed Improvements in the Federal Statistical Programs for City Planning Purposes HENRY COHEN, *Department of City Planning, New York City.*

The recent proposal by Congressman J Arthur Younger (Rep Calif) to set up a federal Department of Urbiculture or Urban Affairs requires serious consideration The paper traces the history and evolution of our governmental structure, with particular reference to the establishment of major Cabinet-rank departments, and shows how slow the country has been in according full recognition to major functions.

City planners are concerned with the physical planning of urban communities containing in excess of half of the nation's population, in which the real estate assets probably equal $400,000,000,000, in which probably over $200,000,000,000 of the national income was produced in 1953, and in which over $9,000,000,000 was spent in 1953 on residential and non-residential construction.

The recent report of the Intensive Review Committee, headed by Ralph J. Watkins, recommended two quinquennial Censuses of Agriculture at an estimated cost of $46,000,000 per decade This is considerably more than the costs of the recommended Censuses of Business and Manufactures, which cover mainly urban areas

One of the dimensions of urban life which has been largely ignored and neglected in the federal statistical programs is the physical and spatial basis.

The eighteen largest cities in the country, each with half a million people or more, contained a total of 26,891,295 people in 1950 compared to a total farm population in the nation of only 25,058,000. The urban population concentration of approximately 11,800 persons per square mile is 850 times greater than the average farm density of 13.8 persons per square mile The organization and management of activities which are so heavily concentrated in small areas is an undertaking of the greatest difficulty. This planning and management task can hardly be handled satisfactorily without detailed data showing the characteristics, intensity of concentration, and location of the varied urban activities and functions.

Business and industrial establishments have (a) structural characteristics, (b) are located in a physical environment, and (c) exist in a pattern of spatial relationships to their markets, their supply, servicing, and financing sources, transportation, and competing establishments. These physical factors and relationships, neglected in the federal economic censuses, are important and measurable A man from Mars studying our society from Census sources would be impressed with the tremendous economic activity carried on by millions of people in a spatial setting which apparently has no physical dimensions.

One of the other major shortcomings in the federal economic censuses (Business and Manufactures) has been the failure to report information for local areas in large cities Though there are separate reports for urban places of 10,000 or more population, Manhattan, for example, with 1,954,000 resident population and 2,466,000 workers employed on the island, is provided with no separate reporting, as a matter of course, for any of its economic districts, e g , the garment district, the port areas, etc

Southern Regional Graduate Summer Sessions in Statistics. GERTRUDE M COX, *Institute of Statistics, Raleigh, North Carolina.*

The Southern Regional Education Board is a public educational agency created in 1948 by the Southern Regional Education Compact among the 14 southern states and is dedicated to the improvement of graduate and professional education in the region A general coordinated statistics program is being developed in the southern region with the following objectives (1) to promote the use of efficient statistical techniques in all fields of research, (2) to provide assistance in planning surveys and in designing experiments, (3) to develop a higher quality and greater availability of teaching, research and consulting service in statistics, and (4) to advance statistics through the discovery of new techniques by theoretical investigations

At the first conference, the group considered the need of existing statistical personnel for additional training, the possibilities of coordinating statistics curricula, the status of consulting services and of contract and cooperative research, and the basis for a regional program in statistics

An Advisory Commission on Statistics was appointed and met April 19–20, 1953 The Advisory Commission made plans to initiate a summer session program It was proposed that those universities having the facilities consolidate their efforts by holding a joint series of summer sessions of six weeks' duration each The first "Southern Regional Graduate Summer Session in Statistics" was held June 9 to July 17, 1954 at Virginia Polytechnic Institute

Graduate students made up two-thirds of those with university affiliation. The universities provided 80% of the total group, government had 14% and the remaining 6% were from industry As planned, the session was of greatest service to university people

Optimum Allocation in Two-Stage Cluster Sampling using Call-Backs on Non-Respondents JOHN E DOWD, *Cornell University*

A model developed by Deming (*Jour. American Stat Assn* , 48, 1953) which allows the calculation of the variance of response and bias of non-response when a simple random sample is drawn, is extended to the case of two-stage cluster sampling

It is assumed that each individual in the population has a certain probability of being interviewed successfully, and that for convenience of identification and computation the individuals' probabilities may be grouped into 5 mutually exclusive classes designated by probabilities, $\pi_i = i/4$ ($i = 0, 1, 2, 3\ 4$).

It is also assumed that the proportion of individuals falling into each response class will vary from cluster to cluster, the proportions being in some way related to other cluster characteristics (e g., income, average rent paid, etc)

Expressions are then developed for the sample mean, expected value and variance of the sample mean, and the bias of non-response, for both the initial attempt and the combination of subsequent call-backs with the initial attempt

By assuming a cost function arising from various components of travel and interviewing costs, an investigation is made as to the optimum manner of allocating resources among clusters in the initial and subsequent attempts at obtaining interviews The allocation of resources is said to be optimum if it produces a minimum mean square error subject to fixed costs

The mean square error of the sample mean here refers to the variance of the sample mean plus the squared bias of non-response.

An expression is also developed for the expected value and variance of the estimate of the population mean when the Politz-Simmons method (*Jour. American Stat Assn.* 44, 1949) for eliminating call-backs is used

For a certain range of values of the assumed cost function and certain other population constants the relative precision of the call-backs scheme and the Politz-Simmons scheme are compared

Factorization of Ethnographic Data. HAROLD E DRIVER and KARL F. SCHUESSLER.

Although indexes of interrelationships among ethnographic variables have been arranged in square matrices since Boas wrote in 1895, clusterings of variables were determined only by inspection until very recently. In 1954 Clements (*American Anthropologist*, 56, 1954) published the first cluster analysis of a set of intercorrelations among cultural groups.

Factor analysis has not been previously applied to ethnographic data, although it has been increasingly employed in physical anthropology in the last 15 years. The present employment of the centroid method of factorization is experimental Factor analysis is regarded by some authorities as more economical than cluster analysis in that fewer factors are needed.

Using the same data (on 16 tribes in northwestern California) as that employed by Clements, the authors found that a first factor, without rotation, is positive for all tribes and accounts for 76% of the total communality. This common influence which appears to dominate all others in the area may be called *Northwest California Culture* Factor II appears less powerful, the two highest loadings being negative and referring to two adjacent tribes in the area transitional between northwest California and central California The highest positive loading is that of the Hupa tribe located in the heart of northwest California We may label this factor *Hupa vs Transitional* The third and fourth factors, by order of extraction, have loadings of about the same magnitude as factor II, and are tentatively designated as *Central California vs Transitional and Yurok-Karok*, and *Coast Yuki vs Hupa-Van Duzen-Sinkyone 1*.

The first two factors are completely in accord with the accepted ethnological picture as outlined by Kroeber and others Northwest California Culture is obviously the dominant one in the area as a whole, and the contrast between Northwest and Transitional is equally unquestioned The distinctiveness of Central California in the third factor conforms to preconceptions, but the similarity of Yurok-Karok with Transitional in the negative loadings seems strange Similarly the isolation of Coast Yuki in the fourth factor is not too difficult to envisage, but the grouping of Hupa with Van Duzen and Sinkyone 1 is unorthodox

Although the application of factor analysis has not provided a simple "explanation" of the similarities and differences among the tribes studied, it has questioned older interpretations and suggested lines around which new ones might be formed

Social and Economic Characteristics of the Population in the United States Directly Dependent on Agriculture LOUIS J DUCOFF, *U S Department of Agriculture*

This paper analyzes some results of a special project which collated for a sample of farm-operator families and households in the United States information from the 1950 Censuses of Agriculture, Population and Housing Three categories of degree of dependence on agriculture were used in this analysis for farm operator households (1) the wholly dependent on agriculture, (2) the partly dependent with agriculture as the major source of the family's income and (3) the partly dependent with nonagriculture as the major source Only about two million farm-operator families, or 38 per cent of the total, fell in the category "wholly dependent on agriculture " The proportions of farm operator families in categories (2) and (3) were 27 per cent and 30 per cent, respectively

It is found that even among "commercial" farms, only half of the operator families were completely dependent on agriculture An additional 35 per cent of the "commercial" farms were partly dependent with agriculture as the major source of the farm operator's family income Among the "noncommercial" farms, which make up 29 per cent of all the farms in the United States, 80 per cent had nonagriculture as the major source of income

The age-sex composition of the population in farm-operator households and the average educational level of farm operators are fairly similar among the three categories of degree of dependence on agriculture The nonwhite population of farm-operator households was more than proportionately represented among the wholly dependent on agriculture Fertility ratios were lowest for farm-operator families that were mainly dependent on nonagricultural incomes The occupational composition of the employed population in each of the three categories shows sharp contrasts and suggests that the classifications used in the analysis achieve with reasonable adequacy an identification of the population wholly or primarily dependent on agriculture

Simultaneous Confidence Intervals Derived From Multiple Range and Multiple F Tests D B DUNCAN and R G. BONNER.

Two new methods are proposed for estimating confidence intervals for comparisons among n means $\mu_1, \cdots, \mu_n$. The first is for differences between single means and is based on the new multiple

range test (Duncan, *Biometrics*, 1955, to be published). The second is for all comparisons of the form $\sum k_i\mu_i$, where $k_1, \cdots, k_n$ are arbitrary constants such that $\sum k_i = 0$, and is based on the multiple comparisons test, Duncan (*Va. Jour. of Sci.*, 1951). The new intervals are similar respectively to the "allowances" proposed by Tukey (1951) and "contrasts" proposed by Scheffé (*Biometrika*, 1953) but employ a new principle. A *p-mean joint confidence coefficient* is defined for every subset of p-means $p=2, \cdots, n$ as the probability of being correct about all intervals involving those p means. A set of intervals with confidence coefficient β, is taken to be a set for which the p-mean joint confidence coefficients are at least β^{p-1}, $p=2, \cdots, n$, the latter values being termed *joint confidence coefficients based on degrees of freedom*. The use of these special values gives the new methods uniformly shorter intervals than those of the comparable procedures and their appropriateness is discussed. Tables for the new methods are provided in the references, (Duncan 1951, 1955)

An Analysis of Agriculture-Manufacturing Differentials in Service Income per Gainful Worker, by State, 1889–1899 to 1929–1939. RICHARD EASTERLIN, *University of Pennsylvania*

This paper is concerned with the relative difference between service income per employee in manufacturing and net income from farming of persons engaged in agriculture in 1889–99 and 1929–39. Drawing principally on census data, an analysis is presented of levels and trends in this income differential by state and the bearing of these state income differentials on that for the country as a whole. The principal conclusions are as follows.

In all states average income in manufacturing exceeded that in agriculture in 1889–99 and 1929–39, though usually not by as much as for the country as a whole. The size of the relative difference varied considerably among states and regions. A large difference was generally due to a relatively low average farm income, not a relatively high average salary-wage. There is a faint suggestion that the size of the income differential varied inversely with the degree of economic development of a state, and that relatively large changes in the proportion of manufacturing to agricultural employment were associated with relatively high income differentials.

In most states the relative income differential narrowed during the period and in some substantially, but the differential for the country as a whole declined only slightly. The contrast between the state and countrywide movements is due chiefly to the large weight exercised by the relatively small number of states in which the differential narrowed least or actually widened, and the absence of a shift in the distribution of agricultural employment toward high farm income states.

Concepts Employed in Labor Force Measurements and Uses of Labor Force Data. A. ROSS ECKLER, GERTRUDE BANCROFT and ROBERT PEARL, *Bureau of the Census*

The uses of labor force statistics are so varied that the question of concepts is bound to be a controversial one. Data are needed for measurement of current changes in economic activity, for general manpower analysis, for research into long term trends, for the study of special problem groups, and for many other purposes. With the help of various advisory committees, the Census Bureau has tried to meet these often conflicting demands by providing detailed information on the present major categories and by undertaking experimental work to measure various problem groups whose classification is under discussion from time to time. Work has also been done to estimate the effect of certain rules such as the exclusion of children under 14 years from the estimates, and the selection of a single calendar week as the time reference for the data.

Much of the discussion of concepts has centered on the classification of persons as employed who did any work during the week, or who had jobs from which they were absent and were not seeking other jobs. Special surveys of part-time workers (those with less than 35 hours of work during the week) have furnished a measure of part-time employment for economic reasons that has proved a useful supplement to the statistics on total unemployment.

Because job attachments of persons not actually at work are in some cases fairly tenuous, it is argued that some categories should be classified as unemployed or not in the labor force rather than as employed. A recent study of the duration of absence from jobs has thrown light on the strength of job attachments. About 75 per cent of the persons absent from their jobs expected to be back at work within 30 days of the start of their absence and therefore could be regarded as having a fairly strong claim to a job. On the other hand, among those not working because of illness or bad weather, longer absences were more common. These findings have indicated the need for some sharpening of the concept.

Another problem area has been the identification as unemployed of persons who are on the border line of the labor force and who may be incorrectly classified as not in the labor force. A number of experiments designed to measure this group have been conducted. They show that perhaps 300,000 to 500,000 persons, largely housewives and teen-age boys and girls, may be in this category.

The experimental program has also included efforts to reconcile the employment data with those from establishment reports. Information collected on dual job holding, on unpaid absences, and on other factors has proved useful, but has by no means explained all the differences. Apparently conceptual differences are not the only factors causing discrepancies between the two types of data. Sampling

and measurement techniques must also be examined. As a result, the authors found that the use of statistical methods in industry have progressed very well and farther than believed. There is an interest, to the extent that 66% of those not using statistical methods did want a summary of the answers, and that 63% of the questionnaires were returned, even in this unsponsored inquiry of a stranger. The kinds of use indicate a growing interest in the more complex methods, and, finally, that the staff workers, many of whom wrote letters in addition to the formal answers, are more aware of the needs than their superiors, who possibly received their schooling before the teaching of the statistical method became widespread.

Multiple Regression with Missing Observations among the Independent Variables G L EDGETT, *Virginia Polytechnic Institute*

A sample with observations missing is obtained from a trivariate normal population Equations for finding the maximum-likelihood estimators of means, variances and covariances have been obtained. Under certain assumptions explicit values of these estimators are obtained. Finally, the problem of finding maximum likelihood estimators of the regression coefficients of x_1 on $x_2, x_3, \cdots, x_n$ holding the latter constant is discussed

Data Allocation in Observing a Quadratic Relation. A DE LA GARZA, *Carbide and Carbon Chemicals Company.*

Data are taken in a pilot plant to study the relation between a quality characteristic $\hat{y}$ and a controlled variable x It is supposed that $\hat{y}(x)=\alpha+\beta x+\gamma x^2$ suitably describes the relation, α, β, and γ being unknown constants, and that the observations of $\hat{y}$ are uncorrelated and have equal variance Due to cost restrictions, only N points (x_i, y_i), $i=1(1)N$, are permitted Due to equipment restrictions, these points must be taken in a specified x-range, x_L to x_H

Let $Y(x)=a+bx+cx^2$ be the least squares estimate of $\hat{y}(x)$, a, b, and c being the least squares estimates of α, β, and γ The paper discussed the problem of how the permissible N observations y_i, $i=1(1)N$, should be spaced in the specified range, x_L to x_H, for the following purposes (1) to minimise the maximum variance of $Y(x)$ for any x inside the range x_L, x_H, (2) to minimise the variance of $Y(x)$ for a given $x=\xi$ outside the range x_L, x_H, (3) to minimize the variance of the estimate c of γ. (1) and (2) provide optimum interpolation and extrapolation (3) provides an optimum test for the hypothesis of a quadratic relation versus the hypothesis of a linear relation

For (1), $N/3$ observations are located at x_L, $(x_L+x_H)/2$, and x_H For (2), the N observations are distributed at x_L, $(x_L+x_H)/2$, and x_H in proportions depending on the extrapolation point ξ For $\xi \gg x_H$ or $\xi \ll x_L$, $N/4$ observations are located at x_L, $N/2$ at $(x_L+x_H)/2$, and $N/4$ at x_H For (3), $N/4$ observations are located at x_L, $N/2$ at $(x_L+x_H)/2$, and $N/4$ at x_H

For the interpolation problem, it was shown that the N observations may be reasonably located at more than three x-locations with little increase in the maximum variance of $Y(x)$ in the range x_L, x_H Applications to minimizing costs of such experiments were made

Migration and Occupational Mobility in a Moderate-Sized Pennsylvania Community. SIDNEY GOLDSTEIN, *University of Pennsylvania.*

The changing patterns of migration and occupational mobility in Norristown, Pennsylvania, in each decade from 1910 to 1950 have been analysed through data obtained from the integrated use of city directories, vital statistics records, and school records The importance of migration in changing the size and composition of the Norristown labor force has declined markedly since 1910 Whereas in the 1910–1920 decade, the net balance of in-migrants over out-migrants was 213 per 1,000 population, by 1940–1950 this rate was only 35 per 1,000. In the former decade, almost all occupational groups experienced large net gains through migration By 1940–1950, however, several groups lost as a result of migration, and the gains of all others were significantly lower than they were in 1910–1920

On the other hand, during the same forty years the importance of occupational mobility in changing the labor force composition was increasing Of the resident male labor force population of the 1910–1920 decade, 75 9 per cent were in the same occupational group at the beginning and at the end of the decade In the 1940–1950 decade, the proportion of occupationally stable persons was only 65 1 per cent. The decrease in stability from 1910–1920 to 1940–1950 characterised all occupational groups except the semi-skilled for whom the stability rates in both these periods were approximately the same. The direction and range of movement of the mobile segments of the labor force of these two extreme decades has not changed significantly

The analysis of the data on migration and occupational mobility suggests that these two processes serve to complement each other and in so doing, jointly serve to meet the changing needs of the local economy and thereby to effect changes in the labor force structure.

Introductory Course in Applied Statistics for Students with Limited Training in Mathematics. C. H. GOULDEN.

A limited training in mathematics is defined as consisting of high school mathematics with courses in algebra, trigonometry, and analytical geometry, either in a final year at high school or in the first year at a university. Stress is placed on the importance of explaining the role of mathematics in statistics and the extent of the mathematical knowledge required by the student. Certain elementary mathematical ideas are taught at the beginning of the course, chiefly simple probability, permutations and combinations, and simple algebraic functions.

Emphasis is placed on the necessity for concrete rather than abstract thinking throughout the course. Examples are given after a thorough drilling on the calculating machine and these are discussed in lectures after the student has become acquainted with the problem. The very difficult point of how to deal with the idea of continuous frequency distributions is handled by studying the distributions of attributes in the laboratory, leading from these by logical argument to the concept of the continuous distribution. The logic of statistics is given an important place and particular stress is placed on the logic of the test of significance.

A brief outline of the course is given together with a list of the text books and references required.

Variance Heterogeneity in a Randomized Block Design. FRANKLIN GRAYBILL, *Oklahoma A. and M. College*

Consider the linear model given by

$$y_{ij} = \mu + t_i + b_j + e_{ij} \qquad i = 1, 2, \cdots n, \; j = 1, 2, \cdots m, \; m > n$$

where y_{ij} is the observation, μ is a general constant, t_i is the treatment effect, b_j is the block effect, and e_{ij} is a random error from a normal frequency function with the following properties: (E denotes mathematical expectation) (a) $Ee_{ij} = 0$, (b) $Ee_{ij}^2 = \sigma_i^2$, (c) $E(e_{pj}e_{qj}) = \sigma_{pq}^2$, (d) $Ee_{pj}e_{qk} = 0$ if $j \neq k$. That is to say, the variance of the errors may differ from treatment to treatment, correlations of the errors are permitted within the blocks, and the errors must be independent from block to block.

In this paper, it is shown that an exact test of significance of the hypothesis $t_1 = t_2 = \cdots = t_n$ can be made by using an extension (due to P. L. Hsu) of Hotelling's T^2 test.

The 2^3 Factorial in a Latinized Rectangular Lattice Design. BOYD HARSHBARGER, *Virginia Polytechnic Institute*

Tests in pairs can be useful to evaluate the effects of treatments or combinations of treatments on the propellant as they effect the missile in flight and at the same time eliminate the effects of atmospheric conditions. Data for such tests will, in general, be secured under several restrictions, such as different elevations, different temperatures, and variable atmospheric conditions. One of the chief interests may be in linear combinations of the treatments or the so-called factorial effects including the interaction of treatments. If the problem is one in which the cost of the tests is important, a very small efficient test is desirable. This property of efficiency is usually found in factorials, and it only remains to fit the factorial in a design that will have the desired restrictions and adjustments. This paper develops a factorial in a Latinized rectangular lattice design so that, (1) the effects of incomplete blocks can be removed from the treatment effects, factorial effects, and the interaction of row and replicate effects, (2) estimates can be provided for the adjusted treatments, (3) mutually independent estimates can be made for the factorial effects, (4) tests of significance can be provided for (a) adjusted treatment effects, (b) adjusted independent factorial effects, (c) adjusted interaction of row and replication effects, (d) row effects, and (e) replication effects.

Error Rates and Sample Sizes in Multiple Comparisons. H. LEON HARTER, *Wright-Patterson Air Force Base.*

A study is made of the error rates, α and β, and their relation to sample size, N, for three test procedures that have been proposed for multiple comparisons. These three are the least significant difference (LSD) test, Tukey's studentized range test, and Fisher's test. Each method can be used for setting $100(1-\alpha)\%$ confidence limits or for making significance tests at level of significance α. In either case one makes C_2^m statements of comparison in an experiment involving m means. Suppose one conducts a series of such experiments. For the LSD test, the probability α_L of an error of Type I is the expected proportion of statements which are wrong. The corresponding probability for Tukey's range test, α_W, is the expected proportion of experiments with one or more wrong statements. For Fisher's test, α_E is the expected number of wrong statements per experiment. A table gives the values of α_L corresponding to $\alpha_W = 0.05$, 0.01 for various combinations of m and N, also corresponding to $\alpha_E = 0.05$, 0.01 for various values of m. Another table compares β_L and β_W for $\alpha_L = \alpha_W = 0.05$, 0.01 for various combinations of m and N, and for various values of $\delta = |\mu_i - \mu_j|/\sigma$. A third table gives the sample size N necessary to fix both α and β at certain levels for both the LSD test and Tukey's test, for various combinations of m and δ.

The Dun and Bradstreet Surveys of Businessmen's Expectations. MILLARD HASTAY, *National Bureau of Economic Research*

This paper is an attempt to validate the working hypothesis that expectations data make a net contribution to our ability to forecast the future of important economic variables. By a "net contribution" is meant that the expectations show substantial positive association with subsequent experience after allowance is made for the relation of these expectations to other data of potential forecasting value that are or become available at the same time as the expectations Such auxiliary data are supplied by the Dun and Bradstreet Surveys themselves in the form of percentage distributions of the reporting firms according to the quality of their business experiences in the period just closed, and by the Department of Commerce in the form of aggregate time series which represent summations for past quarters of the economic variables to which the expectations of individual firms refer

The findings of the paper, while not strictly conclusive, strongly suggest that in forming their expectations business executives take account of information that is not wholly dependent on current or past values of the variables reported on The expectations thus have forecasting value when used in combination with such pre-existing data

As a test of these findings, the Dun and Bradstreet expectations are used to appraise the timing and progress of the current recession on the basis of statistical relations worked out for the pre-recession period The conclusion is reached that a revival of activity was under way by the second quarter of 1954 and that the recession has been a notably mild one

Dollars and the Dollar Area DONALD F HEATHERINGTON, *National Foreign Trade Council*

Since 1945 the position of the United States, Canada and the dollar area as a whole has been fashioned increasingly by factors and forces outside their direct, independent control Vital policy decisions have been taken repeatedly in response to the apparent dictates of special emergency situations rather than arrived at as essential, studied elements in a long-range program There has been an adjustment to a new, special type of environment, with military aid, military expenditures and foreign subsidies bulking large in the balance of payments Only as the general environment changes may we look for any substantial shift in the structure of either the United States or Canadian balance of payments

The dollar area is an even more loosely knit, informal association than its historic counterpart, the sterling area, and owes its existence to natural growth and the reciprocal flow of trade and investment rather than to any decision or declaration on the part of a central authority or the constituent states

The heart of the dollar area is the United States in that it is the only member whose currency would be considered a full central reserve currency This is an acknowledgment not only of the gold conversion facility afforded international holders, but also of the sheer industrial strength and financial resources of the economy These factors, together with the volume of external trade conducted and the worldwide acceptability of the dollar, have made the United States a natural clearing center and the dollar a currency of settlement

Canada stood for years in an indeterminate position, with one foot in the dollar bloc and the other in the sterling system With the passage of time several forces combined to align the Canadian dollar more with that of the United States, and to fix it as a dollar oriented currency

Since 1945 Canada's only means of divorcing itself from the dollar area would have been through the imposition of the most rigid restrictions and the establishment of discriminatory trade controls, either of which if carried to the necessary extremes, would have wrenched the Canadian economy and placed certain segments at a distinct disadvantage

The United States and Canada, together, constitute the hard core of the dollar area as it stands, forming with Cuba and Panama, the inner circle

On the basis of 1952–1953 trade returns, countries in the outer dollar area rely on the United States and Canada as a joint market and source for well over half of their merchandise exports and imports with the percentages ranging in several instances to above 75 per cent

A preponderant share of the external direct investment in the outer dollar area has come from the United States, and much of this investment has been placed in enterprises directly or indirectly associated with the production of those commodities and products for which a market has been or can be developed in the United States.

The supposed "dollar shortage" has been symptomatic of far deeper difficulties, induced or inherent, than the intimated failure of the United States—or Canada—to place enough dollars abroad to meet every purpose During the past eight years there has been, in fact, not a single "dollar problem," but several problems which have focussed on the dollar, and which, while separate, have had a reciprocal influence. In essence the first of these involved a lack of confidence, the second a lack of production and capital, the third a lack of social stability and acceptance of economic realities

Between the beginning of 1946 and the end of 1953 the United States and Canada jointly have placed approximately $153 billion at the disposal of other economies, giving them command over dollar goods and services to this extent.

In the near-term the dollar supply picture is not apt to undergo any appreciable change, either in amount or direction of flow There is no reason to suspect that the direction of primary outflow of United States dollars will change, since the products and commodities in greatest potential demand are those from within the dollar area and raw material or mineral regions Where these dollars in turn are first spent, however, may be expected to undergo considerable change, as local production facilities and competition alter market structures A cautionary note must be sounded against any expectation of a substantial sustained rise in exports from the dollar area following on the re-establishment of some sort of convertibility to the major European currencies.

There is not only considerable sympathetic understanding but also active interest in the United States with respect to the developmental aspirations and needs of the world However, the process by which development occurs must necessarily involve a more effective marshalling of national savings for internal productive investment purposes and/or a less suspicious—at times hostile—attitude toward private direct investment abroad

Needed Improvements in the Census from the Standpoint of Public Housing Users MORTON HOFFMAN, *Housing Authority of Baltimore City*

The needs of a variety of public and private users of housing census data have been well served by the 1940 and 1950 Censuses of Housing The interests of the public housing analyst differ in degree rather than kind from those of technicians of other federal or local programs of community betterment, and also have much in common with the statistical requirements of mortgage lenders and others concerned with real estate markets Certain additions and modifications in the Housing and Population Censuses would enhance their usefulness for those engaged in public and private housing activities Data on conversions, included in 1940 but omitted in 1950, merit inclusion because of their significance for the study of changing housing patterns, and because of the expectation of declines in quality and size of converted dwelling units The sufficiency of only the renter- and owner-occupancy categories is questioned in view of the growth of "contract ownership" schemes in urban areas.

The adequacy of the "dilapidation" concept used to measure structural condition in the 1950 Census is assessed Because local rehabilitation programs, under the stimulus of the 1954 Housing Act, will be causing a sharp reduction in the number of outside toilets or shared baths in particular areas, the substandardness concept of housing and redevelopment agencies, which is based on the combination of dilapidation and presence of three plumbing facilities, will not be nearly as useful in 1960 as in the past Possible scaling devices in the housing quality indices are mentioned The desirability of coverage of sanitary, safety, and environmental aspects of housing from the standpoint of the newer rehabilitation and conservation programs is brought out, but strong doubts are expressed as to the possibility of collecting highly technical environmental data in a mass Census It is concluded that the Census Bureau must keep abreast of the existing and embryonic programs of community improvement so that its data are aimed at assisting in the conduct of these programs and facilitating an evaluation of their effectiveness

Mobility data on a household as well as area basis would be extremely useful to those concerned with the fields of housing and community planning and organization Mobility by race and tenure should be crossed with income, housing quality, family type and size, age of family head, crowding and doubling up Information on family moves over a three-year rather than a one-year period, and on mobility intentions, would also be of great interest The 1950 Census materials on family-housing relationships are applauded, with recommendations that a nonwhite breakdown and more cross-tabulations be included in the future, and that the important "broken family" category be separated out Income data on single persons should not be merged with that relating to families of two or more persons in the housing and family-housing tabulations Concern with the characteristics of displaced families on the part of newer and older governmental housing programs increases the need for family income data for areas smaller than a census tract It is recommended strongly that an intercensal housing survey be conducted because of the need for current data and the importance of experimentation with new techniques and procedures

The Undergraduate Program in Statistics at Iowa State College PAUL G HOMEYER and D. V HUNTSBERGER

A program leading to a B S degree was started at Iowa State College in July 1947, at which time the Department of Statistics was created The objectives of this program, in addition to offering the undergraduate degree are (1) to teach an all-college course in statistical principles at the junior college level, (2) to offer general courses in statistical theory and methods as well as some specialized courses for students majoring in other subject matter areas, and (3) to attract and prepare exceptional students for graduate work in Statistics The first B S degree in Statistics was awarded in June 1949 and to date twenty-six such degrees have been granted Of these twenty-six graduates, seven are working for or have completed advanced degrees in Statistics In addition to the majors, four students who minored

in the department in their undergraduate program have elected to work for graduate degrees in Statistics. The undergraduate major in Statistics consists of a minimum of thirty credits which normally consists of two quarter sequences in Statistical Methods, the Theory of Statistics, and Processing of Data, plus at least four one quarter courses elected from Business Statistics, Experimental Designs, Survey Designs, Quality Control, Economic Statistics, Psychological Statistics, and Industrial Statistics. Each student must also work in two minor fields to acquire a total of thirty credits. One minor is invariably Mathematics and normally consists of at least twelve hours beyond the Calculus. For a second minor the student is encouraged to elect a subject matter area where Statistics has recognized applications.

Problems have arisen in connection with finding suitable texts and personnel but it is of interest to note that there has been almost no objection to consolidating the teaching of Statistics in a single department With increasing numbers of students certain changes can be made to improve the program. At present the undergraduate majors take some courses primarily intended for graduate students minoring in Statistics Some changes in the present courses and the addition of new courses are being considered

Jewish Demographic Research in the United States. C MORRIS HOROWITZ, *Brooklyn College.*

Although the Jewish people, even during the Biblical period have been census-conscious, today the Jewish population in the United States knows nothing about its demographic characteristics. Unlike the official Canadian statistics, the United States Census Bureau does not collect statistics by ethnic classifications.

When the Jewish community in the United States was in its infancy, welfare efforts were little more than sporadic charitable attempts Today, with a Jewish population in the order of magnitude of five million, planning for welfare, cultural, educational and religious institutions and functions must be based on scientifically prepared demographic data.

Numerous methods and techniques have been employed to estimate the extent—and to a very limited degree, the demographic characteristics—of the Jewish population Through the *Census of Jewish Congregations of the United States Census of Religious Bodies*, conducted every ten years from 1850 through 1890 and from 1906 through 1936, the Census Bureau conducted a census of Jewish congregational membership For the census of 1926 and of 1936, the definition of Jewish congregational membership was so broadened in scope, that for all practical purposes, it became a census of the Jewish population The data was collected by extremely questionable methods and from unreliable sources, and yielded information which was not comparable from one census year to another and which was little more than an aggregate of guesses

The *Yom Kippur or School Attendance Method* of estimating the extent of the Jewish population is based on the premise that Jewish children do not attend public school on Yom Kippur, the most holy day on the Hebrew Calendar It involves a comparison of the registration in the public schools, the number of absences on a "normal" day, and on Yom Kippur The result, modified by some constant, supposedly yields a Jewish child population figure. Based on the assumption that the Jewish age distribution is similar to that of the general population, the extent of the total Jewish population is estimated. This technique involves too many questionable assumptions

The technique, *Interpolation from Census Data*, assumes that the demographic characteristics of an area which is densly populated by Jews are the same as the characteristics of the Jewish people in that area The assumptions are many, and with the almost halting of immigration, and consequently, with an increasing proportion of the Jewish population being native born, this method could not be applied.

The *Master List* method involves the formation of a list from membership rolls of Jewish organizations, fund raising campaigns, etc. A questionnaire is then circulated either among the entire list or a portion of it These lists are biased, and one questions the validity of the assumption that the characteristics of affiliated Jews are the same as those of the unaffiliated ones

The *Jewish Name Method* is based on the preparation of a list of typically Jewish names and on a study of the incidence of these names in the community The preparation of such a list is a subjective task and the list of names in the general community to be studied usually turns out to be a biased one.

The *Matching Technique*, developed by the U S Census Bureau, involves the sending of a list of Jewish names and addresses to the Bureau, which will pull out the corresponding IBM Census cards, and run them off for any information desired The preparation of the original list, and the fact that the names and addresses must be as of the census date, prove to be weaknesses of this technique.

The *Birth Rate* and *Death Rate* techniques are based on the determination of either a birth rate or age specific birth rate, or a death rate, or an age specific death rate Basically these methods revolve around the Jewish Name Method. Not only does this technique involve the fallacies inherent in the latter, but develops limitations of its own.

As a positive approach to this problem, it is suggested that the possibility of the organization of a

Jewish Demographic Laboratory attached to some university, possibly the Yeshivah University in New York, should be investigated; and that the feasability of sampling studies, both on a national and local level, should be studied.

Some Recent Advances in Statistics in the U. S Department of Agriculture. EARL E HOUSEMAN, *U.S.D A.*

A sketch is given of some of the developments in statistical procedures for a few broad areas of activity within the Department of Agriculture including the review and coordination of statistical work, research work on problems of crop and livestock estimating, studies of statistical techniques for economic analyses, and the establishment of a Biometrical Services staff in the Agricultural Research Service Current research on crops, primarily cotton and corn, is directed toward (i) getting a better understanding of the concepts underlying farmers' reports of acreage and yield, (ii) determining how much various factors contribute to differences between reported yields and estimates from sample measurements of the crop in the field just prior to harvest, and (iii) further work on sampling techniques. In the field of econometrics, studies have included work on the possible application of simultaneous equation models, spatial equilibrium models, and linear programming, in addition to improvements in methods of using the more conventional least-square technique.

The Relations between Correlational Expressions of Test Reliability and Variance Ratio Expressions. CYRIL J HOYT.

The equivalence of a number of formulas for obtaining the split-half reliability coefficient was previously shown by Cronbach in a 1951 *Psychometricka* paper It was pointed out in Hoyt's paper that these formulas of Flanagan, Guttman, Mosier, Rulon and Cronbach are individually equal to the ratio of four times the covariance of the test halves to the variance of the total scores, when scores are expressed as deviations from their respective means or $r_{11}=4\sum a'b'/\sum t'^2$ where a' and b' are the half-test scores and t' is the whole test score—all expressed in deviations from their respective means. It was also noted that the formula given by Rulon which expresses the reliability coefficient in terms of the ratio of the variance of differences in half test scores to the variance of the whole test is equivalent to the analysis of variance reliability formula, $r_{11}=MS_I-MS_E/MS_I$, where MS_I is the mean square for individuals and MS_E is the mean square for error If one employs the maximum likelihood estimate as given by Jackson and Ferguson for obtaining the estimate of the product moment correlation coefficient between the scores on two halves of the test and then uses the Brown-Spearman formula for obtaining the reliability coefficient of the whole test, this likewise is algebraically equivalent to all the formulas summarised by Cronbach Applications were made to new data It was noted that the use of the usual formula for the product moment correlation between the half tests in conjunction with the Brown-Spearman formula is lacking in consistency of assumptions since the Brown-Spearman formula does assume equal variance for the half-test scores while the usual product moment correlation does not If one is willing to make the assumption of equal variances of the half-tests for justifying his use of the Brown-Spearman formula consistency would demand that he make this assumption earlier in correlating the half-test scores. This assumption would imply the usefulness of the appropriate maximum likelihood estimate as given by Jackson and Ferguson in Bulletin No 12 University of Toronto Department of Educational Research

Personal Saving in Canada: Direct Estimates 1939–1953. D J R HUMPHREYS, *Bank of Canada, Ottawa*

To date in Canada there has been only one published estimate of total personal saving. This is the "residual" estimate of personal saving arrived at in the National Accounts by deducting estimated personal expenditure from estimated personal disposable income This residual estimate of personal saving has two drawbacks First, a small percentage error in the estimate of income or expenditure may produce a much larger percentage error in the residual estimate of saving and because of the magnitudes involved this error in saving may be quite large in absolute dollar terms Second, the resulting estimate is a single aggregate and no clues are given as to its component items and their movements

The "direct" estimate, using a balance sheet approach, analyzes the changes which have occurred in a selected period in the great variety of assets and liabilities of the personal sector which are classed by the National Accounts as saving The net total of these changes produces a saving estimate which provides some check on the residual estimate of personal saving and at the same time reveals the changes which have been occurring within the aggregate total.

There does not appear to be any systematic pattern of differences between the direct and residual estimates and the study would seem to confirm both in absolute terms and as a percentage of personal disposable income the high level of personal saving in the war and post-war period, as shown by the National Accounts.

The direct estimate has been very successful in providing some insight into movements of the com-

ponent items of personal saving. It has made it possible to study the changes which have occurred in either a single year, or over a 15 year period, in liquid asset holdings, in contractual saving such as life insurance, in investment in residential housing, in investment by farmers and unincorporated business in machinery, equipment and inventories, and in such liabilities as consumer and mortgage debt, bank borrowings and payables

In addition, the direct estimate of saving has also provided a number of valuable by-products In the case of liquid assets the balance sheet approach has produced estimates of total personal sector holdings of liquid assets other than corporate bonds and stocks The direct estimate has been used in a preliminary way in the making of GNP-GNE forecasts and later it may make feasible forecasting of the supply and demand for funds by the personal sector In its present preliminary state it has served as a useful springboard from which to launch new statistical studies, the most interesting of which has been the recent Dominion Bureau of Statistics survey of trusteed pension funds It has also been reintegrated with the National Accounts personal income and expenditure detail to provide a modified form of source and use analysis.

The apparent advantages of the direct estimate of saving should provide incentive for further development Now that the National Accounts is producing quarterly estimates of personal saving, the direct estimate should be moved onto a quarterly basis as soon as possible For interpretative purposes attempts should be made to split down the direct estimate of personal saving into its three major groups of savers, that is farmers, other unincorporated business, and other persons By means of direct surveys some attempt should be made to determine the character of saving and investment by the personal sector, by income group and probably by region Finally, the statistics where feasible should be shifted from a net to a gross basis in order that a money flows analysis along the lines suggested by Messrs Copeland and Brill might be made for the Canadian economy

These developmental problems are no longer ones of integration and interpretation of existing statistics They are problems of developing entirely new series, series which necessitate direct contact with non-Government bodies, with the business and financial communities Success will depend on their co-operation

Statistics Curricula at Purdue University Paul Irick

About twenty-five undergraduate, dual level, and graduate courses in statistics are offered at Purdue each semester to an average total enrollment of four hundred students Approximately one-third of these courses deal with the theory of mathematical statistics and probability while the remaining courses are described as statistical methods courses The theory courses are given in the department of mathematics, as are about one half of the methods courses Other methods courses are given in the departments of agronomy, agricultural economics, economics, forestry, and psychology.

One full year course for undergraduate students serves as an introduction to both theory and methods First and second semester methods courses at the dual level are offered in five different areas of application A committee of statistics instructors from the various departments serves to bring about the coordination of these courses, both with respect to course content and methods of classroom presentation Undergraduate students may major in either mathematical statistics or in statistical methods Either of these curricula specifies that the student must take about forty semester hours from the courses in statistics and closely related subjects Undergraduate students who do not major in statistics may elect to take either the undergraduate course or a first and second semester of statistical methods In the latter case, the student may choose from courses which are either general in their application, or which are planned with special emphasis on applications in agriculture, economics, engineering, or psychology

The Adequacy (or Inadequacy) of Statistics for the Purposes of Forecasting in the Field of Mortgage Investments A L Jackson and A. W Gilbart, *The Equitable Life Assurance Society*

Reviewing available statistics and citing their specific uses, it is contended that statistics pertaining to forces outside the real estate field, for example, government monetary policy, business conditions generally, etc , are not adequate to permit accurate forecasting of trends which are necessary in order to forecast the mortgage or real estate market Actual policy as developed under the New Deal administration and the current Republican administration proves the inability of men to predict or to regulate general economic trends.

Specifically citing the errors in forecasting money rates, mortgage rates, real estate prices, etc , during the years 1952–1953 and 1954, it is concluded that this is one field in which the American Statistical Association can perform one of its greatest services

Perhaps the most important conclusion in this paper is that statistics in the insurance company and related investment fields need to be greatly expanded and revised so that we can know more precisely what the supply and demand for funds is, or will be, so that we will know better how to estimate property income and expenses and measure values. In this connection, specific instances relating to housing, com-

mercial properties and industry are cited. Again, the conclusion is much is to be done in the assembly coordination and dissemination of needed statistics, with the American Statistical Association assuming the lead.

Multivariate Analysis J EDWARD JACKSON and ROBERT H MORRIS.

The need for multivariate analysis in quality control problems is illustrated with several hypothetical bivariate examples An ellipse whose size and shape is determined by the covariance matrix of the two variables is suggested as the appropriate control region. The extension to the p-variate case utilises Hotelling's T^2 statistics and, in the cases of degeneracy or near degeneracy, of Hotelling's method of principal components.

For short range control work certain modifications are necessary which involve very simple computations and yet will yield essentially the same results as a full multivariate system A specific application of these methods is made to a photographic color process.

Mobilization Planning in Canada (Summary) R WARREN JAMES, *Dept of National Defense.*

Historically, planning for industrial mobilization in Canada has been neglected because of geographical proximity to a peaceful neighbor, remoteness from the internal troubles of Europe and the subsidiary position of Canada in the imperial political system In wartime, Canada's troops have traditionally participated heavily in the supply systems of allied countries and responsibility for logistical support has been limited Canadian industrial capacity has been large relative to domestic military needs and the major problem of war production has been to fill external demands for munitions whose magnitude cannot be predicted in advance

The development of an independent civilian procurement agency does not facilitate mobilisation planning in the narrow statistical sense nor does the marked preference for competitive defense contracts Some conflict exists between the optimum development of facilities to meet a future emergency and the goal of minimum cost in defense procurement Since 1950, one major advance in mobilisation planning in Canada has been the creation of the Department of Defense Production which has undertaken mobilisation stockpiling and the provision of new capital facilities for defense contractors and has served as a training ground for industrialists and civil servants in the problems of defense procurement and production

Analytical studies of supply and requirements in Canada are basically handicapped by the close interdependence of Canada and the United States Efficient war production in Canada will often depend on the volume of orders received from the United States or elsewhere Forecasting of such external requirements is not very practical Much attention has been devoted to the elimination of legislative and other barriers to procurement by the armed services of the United States in Canada, particularly the Buy American Act Conversely, Canada's dependence on the United States for end products, materials and critical components means that supply levels will inevitably be influenced by administrative and military decisions in the United States whose quantitative effects cannot be predicted

Because of intimate involvement with the problems of diversity of equipment, Canada has consistently espoused increased military standardization in the North Atlantic Community A great deal has been accomplished in converting many classes of weapons and ammunition in Canada to United States patterns This has been supplemented by Canadian participation in the cataloguing system originated by the United States and now spreading to other NATO countries

The Chemical Outlook JEREMY C JENKS.

During the first six months of 1954 the chemical industry operated at about 75 per cent of capacity Dollar sales were a little behind in the first quarter of the year, but gradually worked up to the point where fourth quarter prospects promise to be satisfactory Inventories, although still a bit higher than entirely desirable in their relation to sales, have declined from the top level of the autumn of 1953 At present with working capital of about $4 8 billion, and the long term debt under half that figure, the financial overall position of the chemical industry is comfortable, disproving last year's fears of a collapse.

Although it seems likely that the 10 per cent rate of expansion made annually during the past thirty years will not be entirely maintained in the immediate future, factors in the industry point to a restoration of a satisfactory level

Total operating earnings before taxes for the last six months were about 18 per cent behind 1953 Approximately 25 per cent of this loss may be attributed to higher depreciation and amortization charges In addition to this, inroads on net earnings that have increased are wages, salaries, and selling costs A sharply mounting outlay lies in the budget for research Also, some plant facilities are operating at lower than economic rates Conversely, the rise in the amount of funds used for research expenditures indicates a determination of the industry to assure its future prosperity

A period of moderately declining corporate taxes would bring a return of the long term favorable performance to the chemical industry. Because of the expiration of the Excess Profits Tax Law, net after taxes for the past six months was perceptibly higher than last year

One fine illustration of the industry constantly fulfilling coming needs is the great growth in plastics. This has become a sizeable business Today plastics are seen in the manufacture of automobiles, television sets, and machinery, as well as for shower curtains, floor tiles, and reinforced articles Among the newer offerings is a fibre carpet which will be both long lasting and easily cleaned It promises to be revolutionary. Several forecasts estimate that by 1957 production of plastics will exceed 4 billion pounds compared with about 3 billion at present.

The Predictive Value of Data on Consumer Attitudes GEORGE KATONA, *University of Michigan*

The research conducted over the past ten years by the Economic Behavior Program of the Survey Research Center demonstrates that changes in consumer motives, attitudes, and expectations exert a significant influence on economic fluctuations Measures of such changes, as obtained through sample interview surveys, serve as indications of the direction of forthcoming changes in spending and saving.

Consumer optimism reached its highest point toward the end of 1952 At that time, in contrast to 1951, the Korean war was considered as having favorable effects on the domestic economy, confidence was wide-spread that the newly elected Eisenhower administration would stimulate business, and people were accustomed to the prevailing price level that had risen in 1950–51. Therefore, many more people than before decided to satisfy their needs for newer and better durable goods.

Toward the end of 1953 American consumers looked to the future with less confidence than a year earlier The truce in Korea as well as news about government economies, decline in production, and increase in unemployment created some pessimism which, however, was held in check by satisfaction with personal financial welfare and with price stability. Consumers expected to purchase durable goods in 1954 at a lower rate than in 1953, but the decline was small

By June 1954 consumer sentiment improved again. Appraisals of the economic situation and prospects were only slightly better than at the beginning of the year, but it was significant that the downward trend of 1953 did not continue One major factor contributing to this development was that a great many people were agreeably "disappointed" insofar as dire predictions about a recession or a worsening of their own financial situation were not fulfilled The improvement in sentiment was strongest among urban upper-income consumers.

Needed Improvements in the U. S Census from the Standpoint of Social Statistics Users: Social Statistics Section's Paper—Neighborhood Improvement ALBERT J KENNEDY, *National Federation of Settlements and Neighborhood Centers*

Neighborhood is that aspect of the community process through which contiguous households relate themselves to one another in support of two primary purposes of (1) individual maintenance, repair and operating base, and (2) social replacement, i.e , procreation and child nurture. Neighborhood becomes statistically visible through the sizes, shapes and orientation of massed households, and the institutions, i e , dwelling structures, shops, churches, schools and other organizations which households call into being.

The neighborhood process has three bio-social stages based on space-time-foot movement intervals of sensory perception and response, designated (1) domiciliary cluster, (2) neighborhood, a child oriented interval, and (3) district, adolescent and adult oriented Each stage is structured by the institutions and associations utilized by the integrated households Neighborhood institutions reflect the character of their sponsoring households Child-bearing and -rearing households structure the normative neighborhood.

Community organization needs for its special purposes statistical aids in identifying and locating in municipality the different types of households in and through which individuals of different age, sex, class and cultural characteristics associate, especially those families actively involved in child bearing and rearing

Census Tract P. Tables 1–7 go far toward providing a satisfactory picture of individual status Additional items covering (1) women as operating housewives and mothers, under occupation, and (2) religious affiliation as an index of culture, are needed

Tables H. 1–10 classify households and families as an associative process They should be further developed, especially the husband-wife group, by number of children under 5, 5–14, 15–19 years of age, and by religious background. This material should be included in Census Tract Tables.

Use of the word "room" as sole measure of dwelling quantity and hence of household functioning, is questioned. A further symbol to characterize the useable space, indoors and out, of one and two and some even larger structures, should be developed The items of household equipment, television and refrigerator, should be supplemented by the even more functional items of garage, automobile and

telephone. Households and families, as above, should be cross classified by dwelling structures and by number of rooms, broken down for Census Tracts.

Neighborhood structure is heavily conditioned by the number and by the orientation of households to each other. Rental housing is becoming quasi-public and public in character, and easily dominated by economic and political forces. This represents a departure from our typical American culture, the effects of which should be questioned. The H. 1–10 tables should be developed to show the location, use, condition and social resources of massed one, two and three room dwellings in super structures and super neighborhoods. This material should be tracted

The present H , C. T. Table 3, should be developed as a modulator between P. and H. Tables to make the Census Tract a workable tool for identifying bio-social processes Chief items in this development are institutions classified by functions and auspices

How Small Can the Sample Be? NATHAN KEYFITZ, *Dominion Bureau of Statistics.*

The question of sample size can only be conveniently discussed for probability samples, and where accuracy of the enumeration process is controlled and parameters to be estimated are clearly defined. Surveys such as those of consumer intentions discussed by Katona seem to fulfill these requirements.

The sample required for any given purpose will be smaller where an effective stratification can be found, if auxiliary data correlated with the object of the survey can be brought into use, if the object of the survey may be served by asking for attributes which are fairly widespread, ideally which apply to 50 per cent of respondents, if "oversampling" can be used in appropriate strata, and if quantities rather than "yes" or "no" answers can be asked of the respondent. Smaller samples may be used and better interpretation of results obtained on future sales, of washing machines for example, if respondents are classified according to whether they now have a washing machine and how old it is One may thus be led to study cohorts of washing machines and automobiles just as population students deal with cohorts of humans.

An expression which shows the total error in terms of sampling and non-sampling components indicates that beyond a certain point increase of numbers enumerated in the sample is unprofitable as compared with work on the reduction of enumeration error and investigation of biases of optimism or pessimism which affect all respondents equally.

Intentions data based on a given sample would be most useful if they could be fitted into a prediction scheme which included intentions of investors, governments, etc , and made use of known income demand functions A further increase in the effectiveness of the sample would be obtained by including in the prediction model data on past performance of the economy.

Short-Cut Formulas for the Exact Partition of χ^2 in Contingency Tables A. W KIMBALL, *Oak Ridge National Laboratory*

Irwin and Lancaster [*Biometrika*, 36 (1949), 117–34] derived methods for partitioning χ^2 from contingency tables into individual degrees of freedom From an *rxc* table, the partitioning yields $(r-1)x(c-1)$ quantities which asymptotically are independent and have χ^2 distributions with one degree of freedom. Furthermore, the single degree of freedom χ^2's sum exactly to the χ^2 computed from the complete table The computations, as outlined by these authors, involve the construction and subsequent multiplication of three *rxc* matrices. In this paper, the computations have been greatly simplified. Each single degree of freedom χ^2 is given by a formula which is a function only of the observed frequencies and which closely resembles the familiar short-cut formula for a four-fold table. A general expression is given which permits the construction of such formulas for contingency tables of any order. The method is applied to some experiments which compare the effects of X and beta radiation on mitotic rates in grasshopper neuroblasts (*Chortophaga viridifasciata*)

The Effective Application of Statistical Methods Throughout an Industrial Organization. E. P. KING, *Eli Lilly and Co*

This paper is concerned with non-technical problems of the industrial statistician. Problems which arise from the interrelations of statistician, administrator, and client are discussed. In particular, the list includes 1) determining a statistician's appropriate level of responsibility in an industrial organization, 2) effectively communicating the results of applied statistics to top management, 3) dealing with over-optimistic and over-pessimistic clients, 4) determining the statistician's proper role in long-term co-operative work, and 5) communicating effectively with clients.

The general position is taken that none of these problems has a unique solution. The adequacy of the solution obtained in any particular instance appears to hinge on the capabilities of the individual statistician and how he utilizes them For this reason, most of the suggestions made in this paper are directed toward the statistician, rather than toward the administrator or the client.

Some Sampling Efficiencies of a Double Sampling Device LESLIE KISH, *University of Michigan*.

When listing addresses within sample blocks for surveys the field interviewer hastily assigns economic ratings of L, M and H (for low, medium and high) to dwelling units The means and the standard deviations differ greatly among these strata with regard to social and economic characteristics. E g , the means (and standard deviations) of the liquid assets holdings of spending units in the L, M and H strata are respectively \$660 (\$1,700), \$1,800 (\$3,400) and \$5,800 (\$10,000). The strata comprise about 33%, 38% and 7% respectively of the dwellings, 22% are without rating, chiefly in the open country. These strata may be used for the allocation of different sampling rates for survey interviews, with the aim of increasing precision Tables of the results of computations are given for many items from a recent Survey of Consumer Finances, showing the precision per interview for several allocation schemes. Through greater sampling rates for the "higher" strata, gains in precision up to 50% or more (as against proportionate sampling) may be had for some estimates, such as the mean income, mean liquid assets and the characteristics of the higher economic strata However, this "loading" results in small to moderate (5% to 20%) losses in precision for other items, such as the estimates of many proportions. In the appendix, derivations are given for the variances of estimates of the *total* and of the *mean* of a characteristic for members of a *subclass* obtained from a stratified random sample These are respectively and approximately: $\sum^R(N_h^2/n_h[\overline{M}_h\sigma_{Yh}^2+\overline{M}_h(1-\overline{M}_h)\overline{Y}_h^2]$ and $(1/M^2)\sum^R(N_h^2/n_h)[\overline{M}_h\sigma_{Yh}^2+\overline{M}_h(1-\overline{M}_h)(\overline{Y}_h-\overline{Y})^2]$, where M is the number of subclass members in the population, there are M_h members among the N_h elements in the hth stratum; n_h is the size of sample, $\overline{Y}_h$ is the mean for a characteristic and σ_{Yh}^2 is its variance among the subclass members in the stratum

The Effects of Bimodality and of Skewness in a Population on the Distribution of "t" J C LAYMAN and R A. BRADLEY, *Virginia Polytechnic Institute*

The small sample test, known as the "t" test, for hypotheses on the mean of a normal population developed by "Student" and modified and extended by R A Fisher is still of outstanding importance in statistics For the strict validity of the test it is required that the sample be of independent observations from a single Gaussian distribution In practice these conditions are never more than approximately met. R A Bradley, in 1952, presented a general mathematical approach to obtaining approximations to the distribution of "t" for specified non-normal populations Examples developed were restricted to symmetrically distributed populations In the present paper, the effects of sampling from a mixed population with density function $f(x)=a_1\Phi_1+a_2\Phi_2$, where $a_1+a_2=1$, a_1, $a_2\geq 0$ and Φ_1, Φ_2 are normal or Gaussian densities with possibly different means and variances were partially investigated This density is of interest since, with changes in a_1, a_2 and the means or variances of the Φ's, it may be used to represent populations which are bimodal or which have varying degrees of skewness Two special forms of the general density above were selected and studied in some detail for samples of size two The results obtained from the study of the two density functions considered indicate that the probabilities that t exceed preassigned values to differ only slightly from the corresponding probabilities for the normal density Thus, these results substantiate the opinion that the effects of moderate departures from normality may not be serious in the analysis of variance of sample means

The Development of Census Tract Cities in Canada and the Dominion Bureau of Statistics Census Tract Program. O A LEMIEUX, *Dominion Bureau of Statistics*

In Canada, 14 of the principal cities are divided into Census tracts Two, Winnipeg and Vancouver were tracted prior to the 1941 Census and the remainder between 1941 and 1951

The tabulation program was planned to give Census tract statistics for the various characteristics of the population such as sex, age groups, marital status, origin, official language, years of schooling, labor force, class of workers, occupation groups by sex and earnings of wage-earners Statistics were also tabulated for households showing the number of persons and families per household and families showing children by age groups Finally, statistics were tabulated for dwellings showing type, tenure, years of occupancy and dwellings showing various types of facilities

Many requests were received for these statistics by persons and organizations of different types

The inclusion of census tract tabulations in the over-all tabulation program is costly in time and money It is, therefore, necessary for the Bureau to know how useful these statistics are in order to justify the expenditure.

A Designed Experiment in Engineering. FRED C LEONE, *Case Institute of Technology*

A single factor experiment is noted and its linear model is presented This is followed by an evolution to two and more factors and the inclusion of replication Interaction is briefly discussed

A 3^4 factorial experiment is then presented and analyzed The object of the investigation was to study the manner in which the prominent variables in the metal cutting process, and their possible interactions, contribute to the principal subdivision of the energy used in the process The basic assump-

tion in this cutting process is that the total energy required is divided between that required in shearing the chips from the parent work piece and that required in overcoming the friction between the chip and the top of the tool

The variables included in the investigation were the *rake angle* which was controlled between $-5°$ and $+15°$, the *yield strength* of the material being cut, controlled between 100,000 psi and 200,000 psi, the *cutting speed*, controlled between 75 fpm and 125 fpm, the *feed*, controlled between .005 in/rev and 015 in/rev and the *width of cut*, controlled between 060 inch and .150 inch

It was concluded from the analysis, that all of the factors considered here have a significant effect on the energy distribution in the process Most of the two factor interactions proved to be significant but none of the three or four factor interactions were significant

Responsibilities and Organizational Placement of the Statistical Consulting Group. SEBASTIAN B. LITTAUER, *Columbia University.*

Some industrial jobs are assumed to require so highly specialized techniques that people have been assigned to them expressly because of their knowledge of these special techniques. Yet frequently enough this narrowness of approach yields much less in the way of operably fruitful results than was anticipated. Hence it seems much more pertinent in the present discussion to focus upon the place of statistical method in the large organization rather than to start from the assumption that there must be specifically a statistical consulting group

The practical use of statistical method has pervaded so many aspects of modern profit-making and non-profit making enterprise alike that it is more efficient to look upon statistical method as the province of all hands concerned with problem solving rather than as a function of a special consulting group The statistical thinker may be concerned with the rate of production, the prediction of sales, formulation of inventory policy, control of accident frequency, the control of quality, the design of biological experiments, evaluation of military equipment, control of billing errors, and a host of other large or small problems, all in the same organization Efficient functioning of the statistical thinker requires as much familiarity with substantive fields of knowledge as with particular statistical techniques In the most efficient use of statistical method, it is a fundamental aspect of methodology of complex problem solving, and as such, is intrinsic to operational analysis by interdisciplinary groups

Considering large enterprise for the moment, there can be three levels of responsibility, namely, (1) general policy making, (2) translating general policy into operating policy, and (3) implementing specific operating techniques for realization of operating policy The efficient use of statistical method in the light of this organizational framework by means of operations research, operations analysis, and operations engineering groups, is illustrated with examples from industry

Stability of Production Rates as a Determinant of Productivity Levels SEBASTIAN B LITTAUER, *Columbia University.*

Among the many factors which influence industrial productivity levels, the ultimate determinants are in the technical cause system usually referred to as the production process The type of production process, the degree of mechanization, the worker skill, and other forces contributing to the total cause system are, of course, influenced by general economic and other social considerations, as are, in fact, the productivity goals However this be, the degree of attainment of productivity levels is directly dependent upon the technical cause system This system, regardless of the productivity goals aimed at in order to meet economic and other social objectives, can vary so widely and so irregularly as to bring about in practice great departures from these desired productivity levels

A productivity level is a technical objective which can be predicted from a given cause system. In order for prediction to be possible, there must be evidence that a cause system exists, whose parameters, mean and variability, are measurable Otherwise stated, the rate of production of a given work cycle and the daily or hourly, or other useful time unit of production, must be statistically stable If that condition does not exist, there is not a meaningful level of productivity Furthermore, in order to attain aimed at productivity levels, one must be able to predict the outcome of the given productive cause system and therefore one must have a state of statistical control of the production process

The technical procedures entailed in bringing about statistical stability of given productivity rates usually result in raising these rates until an asymptotic "optimum" level is reached This "optimum" level is dependent on many circumstances, dominant among them are, of course, the state of the manufacturing technology, financial forces, and management attitudes. Evidence of the effects of the technical cause system and of the efforts made to bring about stable production rates is presented

Some Problems in the Development of a Synoptic Climatology. THOMAS F MALONE and ROBERT G MILLER, *Massachusetts Institute of Technology*

In an attempt to ascertain the information concerning temperature and rainfall contained in atmospheric circulation patterns, linear operators relating circulation to weather elements have been

derived from climatic data. These operators are able to explain up to 87 per cent of the variance of mean daily temperature in a contemporary sense and up to 84 per cent for a one-day lag. The operators explain 55 per cent of the variance of an areal average of twenty-four hour rainfall. In seeking an explanation for the high predictability, operators were derived for the prediction of circulation patterns. The correlation coefficient between predicted and observed twenty-four hour pressure changes has been computed for several cases of new data and is observed to go as high as 0.9. The derivation of operators of this kind is feasible only through the use of high-speed digital computers. Possible extension to a four-dimensional climatological model is discussed.

The Interactions of Certain Contingency Tables. NATHAN MANTEL, *National Cancer Institute*

The nature of interaction, and its ordinarily correct interpretation and use in the analysis of quantitative data, is discussed.

A correct understanding of the nature of interaction should lead to correct and valid analyses of contingency table data. However, in practice, many analyses are made because they are quick and easy, to meet other desiderata. Examples are analysis by factorial chi-square, or use of the arc sine transformation. For the purpose of finding interactions these are biased methods.

In any particular problem, a correct analysis must flow from an understanding of the problem at hand. Certain types of analysis and transformations which may be useful in particular problems are presented. The possibility of extending these ideas to more complex contingency tables is discussed.

Recent Advances in Government Statistics. HERBERT MARSHALL, *Dominion Bureau of Statistics.*

An important objective of the Dominion Bureau of Statistics in recent years has been to speed up the issuing of various reports. In pursuing this objective, however, care has been taken not to sacrifice essential quality for the sake of earlier publication. While the attitude of the perfectionist may be avoided statistics should reflect adequately the real economic and social trends and processes which they summate. Where statistics are admittedly only approximations their limitations should be clearly stated in publication. In fact it should be general practice to make every effort to specify the degree of accuracy of each statistical series.

Canada's Labour Force Survey was changed from a quarterly to a monthly basis in November 1952, and a timing of release within 4½ weeks of the reference week achieved. Special efforts are being made to study the problems of non-sampling errors.

It has been decided to extend the mark-sense principle to the agricultural schedule to be used in the 1956 census. A mark-sense card was used very successfully in the 1951 population census. Time for completing the 1951 census was cut in half.

Publication of Vital Statistics has been speeded up greatly by an overhaul of the tabulation program undertaken by the Bureau in cooperation with Provincial officials. The tabulation of the results of a nation wide sickness survey is furnishing information on concepts and definitions which will be of great assistance in the development of continuous statistics in this field. New schedules for hospital statistics, both operating and accounting, arising out of Dominion Provincial Conferences, are now in use and will result in much more complete information.

In current agricultural statistics the use of probability sampling techniques is being extended to supplement the data received through *non-random* mail questionnaires.

In the census of industry, questionnaires are being changed over from a production to a shipment basis. Other changes are being made in the annual schedule to make it conform more closely to business accounting thus easing the burden of reporting, speeding of returns, and increasing the consistency of the statistics.

A considerable widening of the Bureau's information on inventories is in progress. In addition to the statistics on unfilled orders in manufacturing industries a series on new orders received each month is in preparation. In retail and wholesale trade new samples have been designed embracing more kinds of business and permitting better regional estimates to be made. An important development in transportation statistics is a pilot survey of road transport which it is hoped will be the start of national surveys. Quarterly statements of the Balance of International Payments now supplement the annual statement. These appear shortly after the end of the quarter. The main National Accounts also are available now on a quarterly basis. Distribution of individual and family incomes, by size, for the year 1951 have been completed and will be published soon. Work on applying the methods used in the construction of the Bureau's index of industrial production (mainly manufacturing and mining) to the other sectors of the economy is far advanced.

Work on an input-output table is progressing. An experimental table for the year 1949 is in course of preparation. Although experimental and relatively modest in scope it is expected to provide a guide for future development in this field, with special reference to the potentialities of the project for purposes of statistical improvement, integration and development.

The Use of Discriminant Functions in the Estimation of the Proportion of Cases in a Given Population. G. E. McCreary, *Cornell University.*

Approximately seventy trichotomous questions which were administered to 612 community people and 78 hospitalized psychoneurotics were being tested for their ability to yield estimates of the proportion of psychoneurotics in the community. Twenty of these questions were chosen as being the most reliable indicators of psychoneurosis.

In the first phase of the analysis we ignored the underlying dimensions being tapped by the test items and assumed that a linear function of the responses could be used to minimize the amount of overlap, i.e., to maximize the discrimination between the two populations. The exact solution of this problem requires the calculation of a 20×20 matrix of the intercorrelations between items For the three point scale of the items the sums of squares and crossproducts reduce to a rather simple formula which can be calculated by cross-tabulating the responses of each item against one another The extension of this formula to a k-point scale can easily be made.

The inversion of the 20×20 matrix in order to solve for the item weights is a considerable task on hand calculators and is even a sizeable task if one has to program it for a semi-electronic calculator. There were indications in the literature that certain approximations could be used in place of this long exact solution without increasing the error by any sizeable amount Eight of these were tried with the same data as in the full discriminant analysis However, none seemed as powerful, with this type of data, in reducing error as previous authors had indicated.

Error was measured in terms of both Type I and Type II errors (false positives and false negatives) at various cutting points The various approximations were compared as to their ability to separate the community mean from the hospital mean relative to the average amount of variation around each mean. The worth of the method was also measured in terms of the ability of the approximations to distinguish between areas with varying percentages of false positives and false negatives The validity of each method of approximation was tested against 64 recheck interviews in which a psychiatrist rated community respondents according to their degree of functioning

We had hoped that some easy calibration would be nearly as error free and valid as the exact method. Hence we could have recommended a simple discrimination method to other investigators in other societies, in other interview situations, with differing items on the same subject. In fact we ourselves would have liked to use a simple calibration for additional data collected in ways differing from the original data on the points noted above.

Although the exact discriminant analysis requires a considerable amount of work it is considerably more error free and valid Of the four methods using intercorrelations between the items which require a medium amount of work the Horst-Smith is best Of the four methods based on the marginal distributions of the items and requiring little work, a weighting scheme developed by Macmillan works fairly well This last method is, of course, not as error free or valid as the Horst-Smith or the exact discriminant analysis If we use equal weights for the items, it is not possible to distinguish community populations which have differing percentages of psychoneurotics.

Non-Parametric Estimation of Survivorship Paul Meier.

A standard problem in life testing and in medical follow-up studies is the estimation of some characteristic of the distribution of survival times (e g, mean, median, or proportion surviving to a given time) from a sample, each member of which is observed only for a limited time.

Certain advantages of non-parametric estimation procedures are noted, and it is shown that for decreasing time interval sub-divisions the familiar "Actuarial" estimators of the proportion surviving approach a common non-parametric limit which is essentially unbiased This estimator is found to be the maximum likelihood solution of the non-parametric estimation problem, and its variance is shown to be well approximated by a formula proposed by Greenwood

Consideration is given to estimation and mean survival which is the characteristic of first importance in many life testing problems Again, the limiting form of the "Actuarial" estimator is essentially unbiased and maximum likelihood, with variance approximately given by a formula suggested by Irwin.

Measures of Industrial Costs in Relation to Industrial Productivity Seymour Melman.

Investigations of variation in industrial productivity levels indicate that the ratio of labor to machinery costs has been a major determinant of productivity changes over time and of differences at single times Accordingly, the problems of measuring labor and machinery costs affect our ability to predict productivity levels.

The cost to management of employing production workers may be estimated by means of job rates and hourly earnings data for single occupations or for an entire work force. Such measures become less useful estimates of the cost of employing industrial workers as non-wage payments are enlarged There-

fore the use of categories like average hourly earnings confer a bias on ratios of man-hour to machine-hour cost.

Machinery cost may be measured in terms of particular machines or by means of price indexes of selected equipment. For such cost comparisons preferred machines are those whose use directly involves labor replacement The significance of labor-machine cost comparisons is altered as labor cost becomes fixed and machine cost more variable in relation to output

Refinement of labor and machinery cost measures will make possible more precise estimates of the portion of variability of industrial labor productivity that must be explained by other factors

Some Problems in Sampling in Air Pollution. GEORGE H MILLY, *Chemical Corps, U. S Army*

In order to put sampling into perspective the total air pollution problem is viewed as a series of processes comprising an operation These processes include pollution generation, transfer and effects The efforts involved in pollution abatement are seen to fall into several modifying processes, viz control of generation, diminution techniques, and influence on the transfer mechanism By consideration of the entire operation it is shown that knowledge of the process of transfer of pollutant is a requirement for effective attack of the broad air pollution problem It is also shown that lack of this knowledge is the primary source of a number of major recurring sampling problems Available atmospheric diffusion theory is identified as the basis for current knowledge of this transfer process and certain aspects of this theory are reviewed The major shortcomings in respect to solving the sampling problems cited are considered and in general arise from the deterministic form of current theory which fails to account for variability of the atmospheric medium and gives no information on variability of the contaminant concentration field. These shortcomings are seen to be deficiencies in fundamental approach, best overcome by application of statistical methods

Some Elementary Aspects of Applied Statistics. S MONRO.

There are three general sources of variability in statistics used for analysis variability of the thing being measured, errors of computation, errors in data This paper is concerned with errors in data

Shewhart's study of the ordered sequence of data is augmented by the studies of digital preference and of greatest common division of first differences between data The binomial and chi-square distributions are used in the tests with appropriate criteria Some useful rules are suggested

Illustrative examples are presented and some general discussion of the causes and importance of data errors is included

Corporate Securities in the Pension Trust Picture ROGER F MURRAY, *Bankers Trust Company, New York*

On broad economic grounds, corporate securities are especially suitable for the investment of industrial pension funds To the extent that such investments contribute to productivity gains, they ease the burden of meeting commitments to retired workers In general, private securities make a more direct and immediate contribution than public issues

Also, the companies which establish pension trusts and the managers of their investments have a familiarity with, and liking for, corporate securities It is not surprising, therefore, to find the bulk of additions to pension trusts being invested in corporate bonds and stocks.

Although regular buying of common stocks by pension trusts is a market factor of some significance, the impact on prices, both actual and potential, has been exaggerated The buying is neither as large nor as concentrated as frequently supposed. Furthermore, it is sober buying for long-term yield which requires a showing that sound values are offered at prevailing prices

Even though pension trust purchases of equities will not inaugurate a "new era" for the stock market, they will continue to be a factor in the healthy broadening of the capital markets Whether pension trust money bears a venture capital or a seasoned equity label, it clearly adds to the volume of available funds for risk taking In the absence of restrictions or other inhibiting influences, the enlarged flow of funds will search out the full range of opportunities for the progress and growth of a dynamic economy

On the Use of the Range Instead of Standard Deviation. GOTTFRIED E NOETHER, *Boston University*

Many standard test and estimation procedures require the computation of the sample standard deviation It is, however, often possible to replace the standard deviation by the much more easily computed sample range without appreciably reducing the accuracy of the method, at least from a practical point of view The purpose of this paper is to discuss various range methods which have been suggested in the statistical literature in connection with problems involving the parameters of one or two normal populations.

It is well-known that the efficiency of a single range estimate decreases rapidly as the number of observations increases. However, this tendency can be counteracted by dividing the sample into subsamples and computing the mean of the various subsample ranges. The author has prepared a table

giving the best possible division into subsamples of equal size for samples containing from 2 to 100 observations This table also provides the necessary constants for computing point estimates and confidence limits for the standard deviation σ of a normal distribution and for carrying out tests of hypotheses concerning σ The same information is given for the corresponding problems involving the mean of a single normal distribution as well as the difference of the means of two normal distributions.

The efficiency as well as the standardized error of these procedures are discussed. It turns out that for many practical purposes the loss in accuracy is negligible and certainly compensated by the greater simplicity and speed of range methods

Comparison and Evaluation of Microchemical Methods C. L Ogg, *Eastern Regional Research Laboratory, Philadelphia*

A series of collaborative studies of microchemical methods have been conducted during the last seven years to determine the most reliable methods for microanalysis and to improve on these methods when possible The methods shown to be best by these studies have been adopted as official procedures by the Association of Official Agricultural Chemists providing the results obtained in collaborative tests were sufficiently accurate and precise both on the intra- and interlaboratory basis.

To determine which method for a given determination produced the best results, two pure compounds of known composition were sent to a group of 20 or 30 microanalysts with the request that they make at least quadruplicate analyses on the samples by the method they normally used and report all data obtained. Each collaborator was also asked to complete and return a form giving uniform and complete information on his procedure.

The mean and variance of the data from each laboratory were calculated and tabulated Interlaboratory precisions obtained by different methods were compared by applying the F test to the variance of the collaborators' mean values Student's t test was used to determine whether or not the difference between the means for two methods was significant Knowing the true value, procedures which introduced constant errors could likewise be ascertained

In addition to determining the best general method, alternate procedures within a method were evaluated using the same statistical tests The data obtained by the general method were subdivided usually into two groups representing alternate steps in the procedure The F and t tests were applied to these two groups of data to determine which, if either, was the better When a significant difference was found the better step was written into a tentative method. If no significant difference appeared, the simpler, more easily applied, or the more commonly used technique was selected.

The tentative procedure thus obtained was tested collaboratively and if both the intra- and interlaboratory accuracy and precision were satisfactory, the method was adopted as an official procedure by the Association.

The fine cooperation of many microanalysts, plus the application of a few simple statistical tests to the data submitted by them has permitted the Association not only to select the best methods presently available but also to improve on existing procedures

The methods for which official procedures have been adopted are carbon and hydrogen, nitrogen by the Kjeldahl procedure, sulfur, chlorine and bromine Methods for nitrogen by the Dumas procedure, and for methoxyl, ethoxyl and acetyl groups are under study

Meteorological Applications of Statistics H Panofsky, *Pennsylvania State College*

The paper describes some of the uses to which meteorologists have put analysis of spectra and cross spectra Applications range from forecasting of diffusion of air pollutants, and airplane design to study of the general circulation of the atmosphere.

Economic Projections by the U S Department of Commerce Louis J Paradiso, *U S. Dept. of Commerce*

Although the Department of Commerce does not publish forecasts of business conditions, the Office of Business Economics publishes jointly with the Securities and Exchange Commission the results of surveys of businessmen's intentions to purchase plant and equipment. In addition, the Department has published on two occasions analyses of prospective markets based on stated assumptions

The first of the markets studies, "Markets After the War" appeared in March 1943 and was widely used by businessmen in planning for the postwar The general conclusion of this analysis was that the techniques employed were adequate in indicating the rough magnitude of the potential over-all postwar increase in economic activity, but were less reliable for the detailed segments.

"Markets After the Defense Expansion," published by the Department in late 1952, was an attempt to appraise business prospects in the period 1953–55, particularly those factors making for a continuation of a high level of activity or those producing some deficiency of demand This involved a careful examination of the prospects for each of the major segments of the gross national product The implications of the defense program were thoroughly explored, and a survey of businessmen's anticipa-

tions was made for 1953–55. The major conclusions of this study were that 1953 would be another year of good business, but that prospects for 1954 were somewhat more uncertain. Actually, the trend of defense expenditures developed differently from that projected and taxes were reduced. Such developments, as the report states, would have important consequences on the course of business. The study was helpful in concluding that the short-run prospects for business were favorable and that plant and equipment outlays would continue high.

A fruitful approach to the problem of forecasting seems to be through surveys of intentions to spend, such as the Department of Commerce-Securities and Exchange Commission annual and quarterly surveys of planned business investment. The annual surveys have correctly anticipated the direction of change of such outlays for all years since 1945 and have closely approximated the magnitude of the change except for 1950 The results have been somewhat less reliable by industry, by company, and by quarters.

In summary, despite the difficulties and shortcomings of the techniques applied to the development of projections, the results have proved to be highly useful to the business community and have been of value to those needing guides as to the general trend of economic activity.

Cross-Spectral Analysis of Time Series. WILLARD J PIERSON, JR and LEO J TICK, *New York University*.

The notion of cross-spectra is explained in terms of the response of a linear system to a random disturbance. The transfer function is, of course, the ratio of the output to input spectrum. The "phase information" is shown to be contained in the cross-spectra, and a phase function is defined as the arc tangent of the ratio of the quadrature and co-spectrum, (which are the imaginary and real part of the cross-spectrum respectively) which reduces to the usual notion of phase for a single frequency spectrum.

The use of these spectra for estimation of the parameters of a linear system is shown Problems of identification arise here

The Process of Learning by Experiment. E W PIKE, *Raytheon Mfg Co.*

The process of learning by experiment is presented as a servo loop whose action is to bring a theory, a mathematical model of reality, into closer conformity with reality. The elements of this process are (1) the theory, (2) the experimental prescription, an exact statement of the operations constituting an experiment related to the theory, (3) the experimental population, the unknown potential results of indefinite repetition of the experimental prescription, (4) the taking of samples from the experimental prescription by carrying out the experimental prescription in reality, one or more times, (5) the inference from the experimental samples to the parameters of the experimental distribution, (6) the comparison of these inferred parameters with those predicted by the theory, (7) the modification of the theory to conform to the result of this comparison A secondary loop of statistical control information goes backwards around element (4)

Statisticians are, in the broadest sense, specialists in this process Their active interest is usually limited to the hypothesis testing in (5) and (6), to the control loop around (4), and to the probabilistic models which are used in many branches of science (physics, genetics, economics).

While all these elements are present in any experiment, the relative emphasis may shift so greatly from science to science that the common process is hard to recognize For example, in agronomy, the type-science of current experimental statistics, the statistical inference (5) and (6) dominates the scene In physics, it is negligible, while the control loop and the probabilistic models hold the center of the stage Other sciences show other patterns.

The historical development of these concepts by Galileo, Bridgeman, Shewhart, Fisher, and others are sketched.

Time Series Problems in Aeronautics HARRY PRESS, *Langley Laboratory*

This paper reviews some recent applications of random process theory to problems in aeronautical engineering. In aeronautics, a number of problems occur in which the behavior of the airplane under the influence of a random type disturbance is of concern As examples of this type, the effects on the airplane of such disturbances as atmospheric turbulence, fluctuating aerodynamic forces associated with buffeting, and the irregularities of runway surfaces will be briefly described These disturbances, in some cases, give rise to severe structural loads and violent airplane motions Because these disturbances can affect the safety, economy, and performance of an aircraft, the successful design and operation of aircraft require the determination and control of the airplane behavior under the influence of these disturbances.

In recent years, progress has been made in the analysis of some of these problems by applications of random process theory and in particular, the techniques of generalized harmonic analysis. The approach generally involved in these studies is to determine the spectral characteristics of the disturbance either experimentally or theoretically and then to apply the results obtained to the calculation of the spectra and time-history characteristics of the airplane response In order to provide a concrete illustra-

tion of the use of this approach, results obtained in a theoretical and experimental study of the behavior of a high-speed missile in flight through rough air are described.

In the applications of random process theory to aeronautical problems, a number of statistical problems are encountered in both the measurement of the disturbance and in the calculation of the airplane response. These problems are frequently made difficult by the complex nature of the disturbance. For example, atmospheric turbulence is a four-dimensional random process (three space and a time dimension) involving the three components of the air velocity vector. As a consequence, the problem of measuring the characteristics of the turbulence requires the determination of the spectra and cross-spectra for the velocity components. Because of these complexities, such measurements are expensive and require efficient experimental designs with particular regard to the sampling variability. A sampling theory for spectra estimation has so far only been developed for the case of a single disturbance of a Gaussian type. The sampling reliability of cross-spectra estimates, however, has as yet not been established even for the Gaussian case. In addition to these sampling problems, there exist other problems concerned with the airplane response calculations. For the simplest case of a Gaussian disturbance and a linear system, the results are straight-forward since the output spectrum may be calculated directly and, in principle, completely specifies the response For non-linear systems and non-Gaussian processes, the problems are further complicated and a need exists for further work in these areas.

Migration as It Facilitates or Retards Occupation and Employer Mobility ALBERT J REISS, JR., *Vanderbilt University.*

The paper describes and analyzes the nature and extent of employer and occupation shifts for migrant men in an urban work force. The data for employer, occupational and residential shifts are for 2,499 sample cases of white males aged 25 to 64 years in the four cities of Chicago, Los Angeles, New Haven and Philadelphia in 1950. Approximately 16 per cent of all men at work in these cities in 1950 were migrants

Migrants as expected, usually changed their employer in their first post-migration job. Persons who migrated for, or as, an employer were surprisingly stable in their attachment to an occupational level as compared with those who changed employers The older the migrant, the greater the difficulty in obtaining work at the pre-migration level, if employers are changed in migration.

Attachment to an employer varies with the specific occupational assignment of the migrant both prior to, and after, migration. Migrant men in all non-manual occupation groups, except sales, were less likely to change employers when moving than were migrant men in all manual occupation groups

There is a moderate degree of stability in attachment to an occupational level when the jobs before and after migration are compared—58 per cent of all migrant moves are stable There is greater risk and uncertainty in making occupation shifts through migration than if one makes occupation shifts without migration, regardless of the age of the migrant, however The older the migrant, the greater is this risk and uncertainty.

Migrants in the higher status non-manual and manual occupations who change occupational leve are more likely to be downward than upward mobile after migration, while migrants in the lower status non-manual and manual occupations are more likely to be upward than downward mobile after migration. Semi-skilled jobs are the single most accessible job to all migrants who make occupation shifts. Migrants in each major occupation group are somewhat more likely to make occupation shifts to related than to distant occupations, except among proprietors

Some Admissible Tag-Recapture Procedures. DOUGLAS S ROBSON, *Cornell University.*

A discrete chance variable X has the probability distribution $p(x, u)$ where u is a positive integer or zero X is a sufficient statistic and the likelihood ratio is monotone in the following sense

(a) if $p(x, u_0) \underset{x}{\equiv} 0$ then $p(x, u) \underset{x}{\equiv} 0$ either for all $u \leqq u_0$ or for all $u \geqq u_0$

(b) $p(x_1, u_0)p(x_0, u_1) < p(x_0, u_0)p(x_1, u_1)$ if and only if both sides are not zero and $x_0 < x_1$, $u_0 < u_1$.

An estimate of u is to be based upon a single observation x The loss due to making the decision $u = j$ when the true state is $u = i$ is $|i - j|$ Only integral valued decisions are allowed, and a decision function δ is defined as $\delta(x) = \{\delta_\alpha(x)\}$, $\alpha = 0, 1, \cdots$, ad inf, where $\delta_\alpha(x) \geqq 0, \sum_\alpha \delta_\alpha(x) = 1$. A decision function δ is admissible if and only if,

(i) for every possible x there exists an integer $\alpha_x \geqq 0$ such that $\delta_{\alpha_x}(x) + \delta_{\alpha_x+1}(x) = 1$
(ii) $\delta_\alpha(x) > 0$ implies $p(x, \alpha) > 0$
(iii) $x < y$ implies $\alpha_x\delta_{\alpha_x}(x) + (\alpha_x + 1)\delta_{\alpha_x+1}(x) \leqq [\alpha_y\delta_{\alpha_y}(y) + (\alpha_y + 1)\delta_{\alpha_y}+1(y)]$

where $[c]$ is the largest integer in c Proof of the necessity of the montonicity of $\delta(x)$ was given by H Rubin in "A complete class of decision procedures for distributions with monotone likelihood ratios," abstract, *Ann Math. Stat.*, 22 (1951), for a more general problem of which this is a special case If both sample space and parameter space are finite then there exists a unique admissible minimax procedure δ^*

which in general is truly randomised and which attains its maximum risk on the parameter points $u_0 \leq u_1 \leq \cdots \leq u_{n+1}$ and has the property $u_i \leq \sum_{\alpha} \alpha \delta^*(x_i) \leq u_{i+1}$, $i=0, 1, \cdots, n$. Numerical examples of some simple ball and urn models are given in which the risk functions of the minimax and classical solutions are computed

Foreign Financing of Canadian Investment in the Post-War Years A. E SAFARIAN and E B CARTY, *Dominion Bureau of Statistics*

Despite very large imports of foreign capital in recent years, the net contribution by non-residents to the savings used for all types of physical investment in Canada has been small Canadian sources of saving from 1946 to 1953 were large enough to finance all but about 4 per cent of net capital formation (after deducting depreciation allowances), and all but about 9 per cent of gross capital formation aggregating more than $33 billion But not all of the savings of Canadians were used to finance investment in Canada Much was invested abroad, particularly through government loans, and there were substantial retirements of debts which had been contracted abroad Canadian sources of saving have directly financed about three-quarters of both net and gross capital formation in Canada since the war, with non-resident sources directly financing the remainder. The Canadian economy actually generated most of the savings necessary to finance this remaining portion of the investment program, but part was used abroad, part represents earnings retained in Canada by non-residents, and part was set aside to replace durable assets owned by non-residents in Canada

These estimates measure only net use of foreign resources and direct foreign financing in relation to capital formation in Canada They are not indicative of whether the underlying investment decisions originated within or outside Canada Nor do they indicate such other aspects of foreign financing as the availability of venture capital and the borrowing of techniques

These findings are in sharp contrast to the extent of foreign financing in previous periods of heavy investment In the period of 1926 to 1930, for example, the contribution of non-residents on balance was about 25 per cent of net capital formation while direct foreign financing of net capital formation (neglecting outflows from Canada) was nearly 50 per cent The corresponding figures for the period 1946 to 1953 were 4 per cent and 24 per cent respectively.

The tempo of foreign financing has increased greatly since 1949 For many years before this Canada had been a net exporter of capital But direct foreign financing of investment in Canada was one-fifth of the total even in 1946 to 1949 In 1950 to 1953 Canada's net use of foreign resources covered more than 10 per cent of capital formation, while direct foreign financing rose to about 25 per cent

The Best Linear Estimates of the Mean and Standard Deviation of Different Populations Obtained from Singly and Doubly Censored Samples A E SARHAN, *University of North Carolina*

The best linear estimates of the mean and standard deviation of the u-shaped, the rectangular, parabolic, triangular and the double exponential distributions from singly and doubly censored samples are worked out for small samples The variances of the estimates and their efficiencies are calculated The behavior of the coefficients for the largest and smallest observations is noticed in every distribution as the number of the unknown observations increases (n is fixed) These coefficients and the relative efficiencies of the estimates are found to form a sequence in these different populations The effect of the tails on the estimates and their variances are also discussed by considering a skewed distribution

Government Securities in the Corporate Pension Trust Picture R. DUANE SAUNDERS.

Since 1949 the Treasury has been receiving, on a confidential basis, reports from bank administrators on the Federal securities held for the account of corporate pension trust funds From this material certain conclusions can be drawn

The number of uninsured pension plans is significantly larger than had been previously estimated For 1951 the accepted estimate was that there were under 2,000 uninsured pension plans, whereas the Treasury survey (which is not all-inclusive) at that time covered 3,400 corporate pension trusts

In the period since 1949 corporate pension trusts reported in the Treasury survey have increased from 1,900 to 4,900, an increase of about 150% The rate of increase, however, has been moderating

There is a seasonal pattern in the asset growth of these corporate pension trusts First quarter contributions are large and then decline in the following quarters Total asset data were only recently requested from the survey respondents and these replies showed an increase in assets of 7% in the first quarter of 1954 and about 4½% in the second quarter On the basis of these reports it is estimated that the total assets of corporate pension trusts were about $10½ billion in June 1954

In general the investment practices of these funds are similar to the patterns evident in other long-term investor groups During the war years the proportion of the funds invested in Governments rose to almost half of total assets Since then the share of Governments has decline to about 25% However, the asset growth has been so marked that the dollar amount invested in Governments since the early postwar years has more than doubled

Characteristically these trust funds hold longer term Government securities. But like other investor groups these portfolios have been declining in their average length of time to maturity. This, in part, is a result of the shortening of the Federal debt itself, but in addition, investment managers have more than doubled their holdings of short-terms as a per cent of total Governments. At the other end of the scale the holdings of savings bonds are large—nearly a third of Government portfolios This was to be expected as the number of trusts are large, the annual limit on purchases has only a minor influence Also savings bonds are especially suitable for these funds

The broader generalisations do hide one of the most striking features of the investment practices of corporate pension trusts. There is an extremely wide diversity in the investment practices of these funds. Banks administering sizeable funds show from 9% to 66% of total assets invested in Governments, that same diversity is evident in all the reports right down through the smaller funds. The appeal of savings bonds as an investment medium also shows this same diversity Although some of this variety may result from the relative newness of many corporate pension trusts, it is more likely that the lack of uniformity in the structure of pension plans themselves is the major reason. Pension plans are in large part specific to the individual corporation and the solution to the pension problems of each firm have been met in a variety of ways Until such time as the structure of the pension plans themselves becomes more uniform, these diverse investment practices will continue.

The Survival of Red Blood Cells. MARVIN A SCHNEIDERMAN, *National Cancer Institute.*

Mathematical models of the survival of transfused red blood cells are considered, and some of their limitations are discussed A chronological development is given showing the emergence through biological experiments of new concepts of the behavior of red blood cells—which follow as a consequence of the simple mathematical models. Treating the population of red blood cells as a stationary population, a short discussion is given showing how birth rates can be found from a consideration of the survivors of a random cross section of the population in which the ages of the individual members are not known An experimental technique, which is very useful from the biologist's point of view, radioactive tagging of the red blood cells, leads to some problems in estimation that have not yet been solved

Developing Work Units in a Child Placing Agency E E SCHWARTZ, *Social Security Administration.*

Units of count used in many social work agencies are not suitable for the purposes frequently made of them The four types of count currently in use in agencies characterised by the use of the case work process are case counts, activity counts, staff counts, and financial data The basic unit, the "case", was originally intended to relate to the individual or family receiving service The practice has developed of attempting to load onto this term the burden of measuring the case workers' responsibility, the volume of work performed by the case worker, and the extent of the services provided Because the "case" is not, in fact, a homogeneous unit of count describing work performed or service delivered, it cannot be related with precision to either the count of personnel providing service to get average work loads or to agency expenditures for service to get costs per unit of service Both of these important averages require measurements which relate the use of agency resources to the services provided A pattern of units of count for social work agencies is set forth to measure input-output relationships through a statistical work measurement procedure A pilot study now in process to test the feasibility of applying work measurement to a child placing agency is described Particular emphasis is placed on the development of work units and problems encountered in their application to programs for placing children for adoption and for foster care.

How Good are Current Statistics for Following Economic Changes? WILLIAM H SHAW, *E I du Pon de Nemours*

The goodness of current statistics for following economic changes is analysed first by discussing how good they need to be, and second by testing the needed goodness against four standards These standards—accuracy, timing, correspondence with market reality, and classification and detail—have to be applied qualitatively because most key statistical series are not susceptible to quantitative tests of adequacy.

Appraisals of goodness using each of the standards are presented for national income and product figures, construction activity estimates, plant and equipment expenditures, the Department of Commerce measures of sales and inventories, the Federal Reserve Board indexes of industrial production, and the major employment estimates Even though the appraisals are slanted deliberately toward realistic needs rather than unattainable ideals, substantial defects or gaps are noted for most of the series

The fact, however, that many improvements in current statistics are needed should not obscure the more important fact that twenty years ago most of the appraised series did not even exist Moreover, significant quality advances have been made during the past five years The producers of the various series have used their modest resources well but for the most part have been unable to get even small

additions to these resources. The problem of getting additional resources is unlikely to be resolved satisfactorily unless economists and business analysts can do a much better job than they have done to date of showing a positive relationship between good statistics and the making of good policy and managerial decisions.

Optimum Sampling in Binomial and Multinomial Populations P. N. SOMMERVILLE, *Virginia Polytechnic Institute.*

Let $\Pi_0, \Pi_1, \cdots, \Pi_k$ represent $k+1$ multinomial populations with unknown parameters $\alpha_{0j}, \alpha_{1j}, \cdots, \alpha_{kj}, j=1, 2, \cdots, p$ We wish to choose the population Π_i for which the function $\sum_{j=1}^{p} g_j\alpha_{ij}$ is greatest. If the sampling must be done in one stage, then results stating the "optimum" amount of sampling that should be done prior to making a choice are given The "optimum" sample size is the one which minimizes the maximum expected loss, taking into account the amount of use to be made of the choice, the cost of sampling and the cost of making a wrong decision.

Quantitative Methods in Physical Anthropology. J N SPUHLER, *University of Michigan*

Quantitative methods have been of some, but quite unequal, use in each of the three general problems of primary interest in physical anthropology during the past 100 years (1) Human evolution, (2) Classification of the living varieties of man, and (3) Comparative growth and development of man. This paper is concerned mainly with the problem of classification

For many human populations of classificatory interest, probability sampling based on available methods would be relatively simple and routine Failure to employ satisfactory sampling methods is one of the major defects in modern anthropometric practice. But in the case of certain populations of anthropological interest, adequate mathematical models are not available. The problem of defining and sampling a local breeding population centering about some point within a large population with area continuity, variable density, and high internal mobility is used to illustrate the need for new statistical models in the study of physical anthropology.

Three methods of estimating the biological distance between two or more human populations using anthropometric data are discussed (1) The seriatim method where differences between means of characters are judged or tested taking the characters one at a time, (2) The Coefficient of Divergence for Multiple Measurements (CD) of Clark, and (3) The D^2 methods of Mahalanobis and Rao

Three difficulties in using the seriatim method are (1) All characters are assumed to have equal classificatory weight, (2) Correlation of characters is ignored, and (3) The probability statements of significance tests are not appropriate measures of biological distance D^2 overcomes all three difficulties D^2 is expensive in computational labor when more than a few populations and characters are classified, but this expense is not large (when machine methods are used) compared to the cost of collecting the original data.

CD and D^2 are identical for zero correlation CD overcomes the first and third difficulty inherent in the seriatim method, but (where $r \neq 0$) not the second Computational labor is relatively small for CD, even for relatively large numbers of populations and characters CD is an approximate method of unknown reliability CD suffers from lack of a significance test (when $r \neq 0$), a difficulty which may be overcome partially by using large samples CD and D^2 are shown to reach identical classificatory results (membership in "clusters") in a study of 8 populations using 8, 9, and 12 characters

An advantage of CD over D^2 is that it may be computed where data are restricted to means—a very common situation in anthropometric reports published during the past 100 years. Where individual data are not available, CD is to be preferred over the seriatim method for estimates of biological distance used in anthropological classification

Unemployment Statistics and Economic Policy Uses CHARLES D STEWART, *U S Department of Labor.*

Section I of the paper deals primarily with the rationale of the concept of unemployment used in the current measurement of the labor force by the U S Bureau of the Census, and its relevance for major purposes of economic policy Section II deals with outstanding questions of measurement other than those of sample design and sampling variability—whether in fact the Census survey succeeds in its purpose of enumerating all persons actually in the labor market with jobs or seeking work. Section III raises, briefly, a more fundamental question whether labor force statistics measuring labor market attachments under existing conditions of demand bring within the scope of measurement all those persons "able, willing, and seeking to work" described by the Congress as within the scope of national economic policy. Section IV considers needs for additional information on unemployment, and concludes that more or less additional data of various kinds are needed depending upon the nature of economic policy—whether largely limited to fiscal and monetary policies directed toward stabilizing or expanding aggregate demand or aimed more directly toward aid to particular groups, industries, or geographic areas.

Analyses of Experiments with Correlated Observations and Heterogeneous Variances. H. C. SWEENY, *Virginia Polytechnic Institute.*

Methods for the analysis of experiments when the usual assumptions of independence between plots and homogeneous variances are not tenable have been investigated, following the method noted by F. Graybill for the analysis of randomized blocks design. Detailed methods for analyzing factorial arrangements and split plot designs are advanced, and the various properties of the tests of significance in relation to the usual analysis of variance have been investigated. Extension to the case of incomplete block designs is underway. The method depends upon a transformation of the data, and an application of Hotelling's T^2 as given by P. L. Hsu to test homogeneity of means.

Problems in the Estimation of the Population Group Subject to Benefit from a Mass Therapeutic Trial. DONOVAN J. THOMPSON, *University of Pittsburgh.*

The estimation of the prevalence of a disease entity or condition in a large population group from survey data frequently must be undertaken in the planning or during the execution of a mass therapeutic trial. Experience gained at the Graduate School of Public Health, University of Pittsburgh in the estimation of prevalence of two chronic diseases, arthritis and heart disease, by three survey methods indicates that major differences exist between methods. The magnitude of the procedural biases (due to the so-called non-sampling errors) of two types of interview surveys is estimated in terms of a "true value" defined operationally by the findings of a physician during a physical examination of the same persons included in the two interview surveys.

Preliminary analyses of the data from two separate studies (arthritis and heart disease) indicate that the "true value" for the prevalence figure is in the range of 2 to 10 times that reported by a household survey employing any responsible adult as the respondent, or a survey employing a randomly selected individual in the household reporting only for himself. In addition, treating the physicians evaluation as the true condition, the household interview was found to give 16.5% false positive reports and 31.2% false negative reports for arthritis or rheumatism, while the corresponding figures for the personal interview of each sampled individual are 17.6% and 24.4%. Similar results were obtained for heart disease. The findings also suggest that the higher prevalence figures obtained from analyzing the reports by a household respondent for his own health condition as against the health condition of other members of the household may well be explained in terms of the higher probability of an ill person being at home and hence more likely to become a respondent on a household survey.

A Topic in Variance Components Analysis. W. A. THOMPSON, JR., *Virginia Polytechnic Institute.*

A lemma is proved which may sometimes be used to find the class of all statistics whose distributions are independent of the nuisance parameters. The least squares model with errors arising from two sources is then discussed, and the lemma is then applied to this case. These results are then specialized to partially balanced incomplete block designs.

Statistics and Educational Research. DAVID V. TIEDEMAN, *Harvard University.*

Consider N people on each of whom is available a series of k observations. This observation matrix of N rows and k columns may be partitioned in a number of ways. Typical educational and psychological problems lead to various partitions and procedures.

Following Bartlett, an attempt was made to show the interrelations among a number of these procedures. The theory of canonical correlation provided the unifying principles. It was recommended that 1. Additional training in the theory of matrix algebra be incorporated in the training of educational researchers, 2. Further training in multivariate statistical techniques be incorporated in the training of educational researchers, 3. An effort be made to find techniques for analyzing the dispersion of multiple variables analogous to available procedures for analyzing the variance of a single quantitative variable, 4. A technique for studying the trace which the points for N people make through time in a T-dimensional space be sought, and 5. Ways and means of obtaining and supporting a large electronic computer for educational research be considered.

The Use of Statistics in the Physical Sciences from the Standpoint of Industry. J. H. TOULOUSE.

The use of "statistical methods in the physical sciences" cannot be characterized by a single measure. The range of use is extreme—from organizations which use statistical methods in every way possible—to those to whom they are virtually unknown.

A small questionnaire was used to sample industrial organizations, and of the 423 sent out 253 were returned, about equally divided between engineers and scientists. A name was chosen on approximately each twelfth page of *American Men of Science*, and each fifteenth page of *Who's Who in Engineering.* The name chosen was that of an individual whose title or description indicated the field of development or research in industry, but not necessarily the manager of a department.

Questions were chosen to indicate interest as well as usage. Answers were generally developed as "Yes" or "No", with some possibility of qualification. The results can be modified by the fact that 84 scientists and 86 engineers failed to return the questionnaire.

Industrial Statistics Needed for Mobilization Planning. WILLIAM TRUPPNER, *Dept. of Commerce*

Experience in both World War II and in the Korean War provides overwhelming evidence that the task of establishing an effective system of production and material controls will take almost a year, if the government has to use the methods developed in the past. Preparation for a future emergency demands that a method for mobilizing our industrial plant be devised for accomplishing this in a much shorter period of time.

Delay in the past has been due primarily to the problems associated with the establishment of production control machinery in the form of orders and regulations to carry out government decisions during a war emergency. Second, delay was occasioned by the inability of the government to measure the impact of those decisions on economic resources.

The development of a production and material control system for use in a future emergency represents a cornerstone of the mobilization preparedness program. The necessity for making the rules effective in a short period of time requires that there be a minimum of change from normal peace-time materials purchasing and production scheduling procedures.

The development of a standby control system for such use is accompanied by the accumulation of statistical measures for translating given program levels into the material tonnage requirements needed to support them. Upon completion of this phase of the preparedness work, it is felt that the two major reasons for past delay in achieving rapid industrial mobilization will have been eliminated.

Some Aspects of Estimation Theory (Preliminary Report). M. C. K. TWEEDIE, *Virginia Polytechnic Institute*

W. L. Stevens (*Biometrika*, 37 (1950), 117–29) has published a method of finding fiducial limits for the parameter of a discontinuous distribution for which the probability of covering the true value can be fixed at a desired level in (0, 1), by utilizing an auxiliary random variable Y which is uniformly distributed over (0, 1). As an alternative to his solution, set up an "acceptability function" (a.f.) $a\delta(\theta|x)$, which ranges from 0 to 1 as θ varies, with the property that the parameter value θ is covered by the chosen estimate set (or confidence interval or fiducial interval), derived from observed data x by a method δ, if the observed value of Y as above is not greater than $a\delta(\theta|x)$. Stevens's result can be used to construct an a.f. Such functions might be used as generalizations of confidence intervals for presenting inferences in estimation. The total probability of covering θ is $A_\delta(\theta|\theta^*)=E_x\{a_\delta(\theta|x)|\theta^*\}$, where θ^* is the true value. Estimation bias may be said to occur if $A_\delta(\theta|\theta^*)>A_\delta(\theta^*|\theta^*)$ for some $\theta\neq\theta^*$. The a.f. derived from Stevens's solution for a binomial distribution is biased but an unbiased modification has been found. Further developments of the general idea are under examination.

Analysis of the Variability of Growth of Filarial Worms. MARY G. WESTBROOK and J. ALLEN SCOTT

A useful criterion of immunity of cotton rats to infections with the filarial worm, *Litomosoides carinii*, is the slower growth of the worms in immune as compared to non-immune hosts. An analysis of the variability of growth in non-immune animals was attempted to provide improved experimental design and possibly reduce the number of parallel control animals needed. Since a separate experimental series to provide data for such an analysis would be prohibitively expensive, measurement of 1583 worms from rats serving as controls to other experiments were utilized.

Pearl-Reed logistic curves gave a satisfactory fit to the means of all worms for each day of age when the goodness of fit was based on the variability of the means about the fitted curve. Confidence bands were desired but impossible to establish because of heterogeneity of variance and the many components of variance involved.

The growth of some free living animals can be studied by successive measurement of the same individuals. These internal parasites, however, are sacrificed when they are recovered and measured. Furthermore, the specific environment in which they were living, i.e., the host, is likewise destroyed. Therefore, growth studies must depend on means of the measurements of groups of worms from a series of hosts. The principal components of variance are then (1) within rat variance, (2) between rat variance, (3) between day variance due to regression and (4) between day variance not due to regression. Even if the data had been derived from an experiment designed for this purpose, lack of orthogonality and heterogeneity of variance would have been inherent. With the data available analysis of variance involving all of these components was clearly impossible. It was possible, however, to show that both the within rat variance and between rat variance were significant. Furthermore, sufficient data were available for one time period to further analyze the between rat variance and show significant effects of the age and sex of the rats.

Fixed, Mixed and Random Models. M. B. WILK and O. KEMPTHORNE, *Iowa State College.*

A procedure is exemplified by which the appropriate linear model for the analysis of a randomized experiment can be derived. A population model is defined from the experimental situation, and a statistical model obtained by imposing the conditions of the design. The physical meaning and the properties of the components of the model are discussed in detail. Using the statistical model, the expectations of the analysis of variance mean squares are obtained and, on the basis of these, proper error terms are specified.

Two examples are considered (i) a general two factor experiment in a completely randomized design, with fixed, mixed and random models included as special cases, (ii) a more complex experiment described by Vaurio and Daniel which has been discussed in detail by Scheffe, for which we contrast our model and analysis with Scheffe's.

Some discussion is given of the advantages and problems associated with the method for model derivation presented.

The Use of Incomplete Block Designs for Factorial Experiments MARVIN ZELEN, *National Bureau of Standards*

The use of factorial designs has now become widely accepted as an efficient way for carrying out experiments involving several factors where each factor is varied at several levels. However, one of the difficulties of carrying out factorial experiments is that many of the block arrangements available are for designs having the same number of levels (i e , p^n series) There are few block arrangements for mixed factorial designs (i.e , factors not all at the same level). The purpose of this paper is to show how incomplete block designs (particularly the partially balanced designs) can be used to obtain designs for mixed factorials systems and in some cases for obtaining fractional replicated designs for mixed systems

An example is given showing the experimental arrangement for a 3×5 factorial design in six blocks of five experimental units which was used to investigate certain techniques involved in the development of printed circuits

STATISTICAL ABSTRACTS

All communications concerning this section should be addressed to the Abstracts Editor, Professor George E. Nicholson, Jr., Chairman of the Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Aitchison, J. A. and Brown, J. A. C., "An estimation problem in quantitative assay," *Biometrika*, 41 (1954), 338

The model proposed by Finney for the quantitative response, u, to a stimulus, x, implies that the variance of u shall be independent of x. The authors have found that for certain economic data this model is unsatisfactory, and that it is more accurate to regard the coefficient of variation of u as constant They therefore propose a new model which uses multiplicative as opposed to additive errors, and show that for this model the variance of u has the property required by their economic data The paper then provides a probit method of estimating the parameters of the new model. There is a worked example WALTER L SMITH, *University of North Carolina.*

Anscombe, F. J., "Fixed sample-size analysis of sequential observations," *Biometrics* 10 (1954), 89–100.

In many cases the investigator sets out to take observations, but has no fixed idea of how many he shall take before stopping He may stop at the end of a certain period of time, or at the expiration of certain funds, or upon some other occasion, which means that his sample size is a random variable. This paper concerns itself with the consequences of analyzing N observations (N random) by conventional methods (for which N is fixed) If the number of observations does not depend upon the observations themselves, then the use of fixed sample size methods introduces no bias or other difficulty. If N does depend upon the values of the observations then several cases must be distinguished (a) First, it is possible to set oneself the rule "take samples until attaining a significant result (by fixed sample size analysis) for the data in hand " This problem is related to non-publication of "negative" results. (b) The experimenter may take one fixed sample of size n, and then a second sample of size n_1 only if the first sample gives a value lying in a normal range (e.g. between $-a$ and a, or positive). (c) The experimenter may take observations sequentially until he obtains a (fixed sample size) confidence interval of preassigned width Errors arising in these three cases are discussed. The author concludes that appreciable error may be suspected when both:

(1) "The number of observations depends on the observations themselves."
(2) "The relative dispersion of the number of observations in repeated sampling is not very small."

Condition (2) would hold in (a) and (b) above, but not in (c) if the preassigned length were quite short). L E MOSES, *Stanford University*

Arbous, A. G. and Sichel, H. S., "New techniques for the analysis of absenteeism data," *Biometrika*, 41 (1954), 77–90.

The authors suggest as a model for absence-proneness a symmetrical bivariate negative binomial distribution For application to industry needs, it was found that accumulating one year's data is necessary before preventive action can be taken on absenteeism These data are halved and by use of correlation techniques the phenomenon of proneness is established. The observed distribution provides estimates that are used for prediction purposes for the following year.

The model may not be the correct one to use, the authors say, but it is sufficiently accurate for practical purposes provided the two consecutive exposure periods do not exceed one year Prediction of the absences of individuals in the second year (before they have actually occurred) is carried out with efficiency which can be established in terms of operating characteristics However, when remedial measures are applied at the end of the first year (before collecting the second year's data) it will not be possible to test the efficiency of that model. Means of assessing the efficiency of the remedial measures taken are not available at present, but the authors indicate possible

lines of investigations along which a solution may be reached. A. R. KHALIL, *North Carolina State College.*

Auble, Donavon, "Extended tables for the Mann-Whitney statistic," *Bulletin of the Institute of Educational Research*, Indiana University, Vol. 1, No. 2 (1953), i–iii and 1–39.

Auble, John Donavon, "Extended tables for the Mann-Whitney statistic and illustrative applications of certain nonparametric tests of significance." *Studies in Education 1952*, Thesis Abstract Series, School of Education, Indiana University, No. 4 (1953), 9–13.

Critical values are given for the Wilcoxon (sometimes called Mann-Whitney) rank order test for equality of the populations from which two samples come All sample size combinations from 1 through 20 are covered One-sided significance levels for which critical values are given are [0 1%], 0 5%, 1%, [2%], 2 5%, [4%], 5% and [10%]. (Bracketed per cents are not given in the second reference) A general discussion of the Wilcoxon (Mann-Whitney) test is given together with (in the first reference) an example of its use The complete distributions computed by the author may be obtained by ordering Document 3912 from ADI Auxiliary Publications, Photoduplication Service, Library of Congress, Washington 25, D C The charge is $4 25 for microfilm or $12 50 for photoprints readable without optical aid. WILLIAM KRUSKAL, *University of Chicago*

Bechhofer, R. E., Dunnett, C. W. and Sobel, M., "A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance," *Biometrika*, 41 (1954), 170–6

Consider k normal populations $\{\pi_i\}$ with unknown means μ_i and variances $\sigma_i^2=a_i\sigma^2$ (σ^2 unknown). This paper is concerned with a two-sample procedure for ranking the populations according to the means with goals the selections of (1) the population with the largest population mean and (2) the populations having the largest, 2nd largest, $\cdots$, smallest population mean. A first sample of a_iN_0 is taken for the i-th population and S_0^2 (unbiased estimate of σ^2) is calculated with $n=N_0\sum_{i=1}^k a_i-k$ degrees of freedom A second sample $(N-N_0)a_i$ is taken such that $N=\max\,[N_0,$ smallest integer $\geq 2S_0^2\,(h_n/\min\,\delta)^2]$ where $\min\,\delta=\min\,(\mu_{(k)}-\mu_{(k-1)})$ for goal (1) and $=\min\,(\mu_{(i)}-\mu_{(i-1)})$ for goal (2) to be detected and h_n is a positive constant involving the goal, n and the probability specified for achieving the goal when $\delta\geq\min\,\delta$. By properly choosing h_n, the solution is to rank the population according to the overall sample means for goal (2) and select the one with the largest overall sample mean for goal (1) Values of h_n have been derived for $k=3$; for $k\geq 4$, the integral remains to be solved for h_n An expression for the expected sample size N has been derived and a particular case represented graphically. Two examples are considered, but in one the first sampling procedure has been substituted by an earlier estimate of σ^2. V. P. SHAH, *North Carolina State College.*

Cadwell, J. H. "The probability integral of the range for samples from a symmetrical unimodal population," *Annals of Mathematical Statistics*, 25 (1954), 803–6.

The author gives an asymptotic expression for the probability integral of the range for samples from a symmetrical unimodal population He calculated the maximum error for two expressions by taking the differences between exact values obtained by evaluating the $p\,d\,f$ and values found by quadrature He found that his second expression gives results of reasonable accuracy. He, then, calculated a table which gives corrections to units in the fourth decimal place to be added to the approximate value for five sample sizes One can interpolate graphically for n and the approximate value if the correction is plotted against the approximate probability on arithmetical probability paper. He also tabulated the percentage points of the range found by quadrature This will help in making preliminary estimates. A E SARHAN, *University of North Carolina.*

Calvin, Lyle D., "Doubly balanced incomplete block designs for experiments in which the treatment effects are correlated," *Biometrics* 10 (1954), 61–8

The mathematical model underlying the design postulates, for every pair of treatments i and j, the existence—and allows the estimation of—an effect arising from the presence of both treatments in the same block, which effect acts oppositely upon the elements given treatments i and j; this correlation parameter is assumed the same in every block

The model is more general than the usual one for balanced incomplete blocks The analysis is simplified by imposing the condition that not only every pair of treat-

ments appears equally often together, but also that every triple of treatments does. This gives the name doubly balanced incomplete blocks.

The construction of such designs is investigated. Designs rarely exist for small r. In organoleptic tests small block size is necessary, but much replication (large r) is ordinarily desirable Usually the number of treatments, p, is less than 20

Several designs are given Others are indexed The analysis is derived and a worked example is given. L. E. MOSES, *Stanford University*

Cox, D. R, "The design of an experiment in which certain treatment arrangements are inadmissible," *Biometrika*, 41 (1954), 287

Suppose that the experimental units of an investigation are arranged in sets of k units, and that the k units within a set have some definite ordering (e g , in time) If systematic differences exist between the serial positions within a set and if there exists an appreciable component of variance between sets, then in a comparison of a number of treatments we would ordinarily use a Latin square or a Youden square However, it may be impracticable to apply treatments in certain orders, and when this is so these standard designs are useless This paper discusses what can be done in such circumstances and, on the basis of the usual assumptions in this field, a method is derived for constructing appropriate designs For a few special cases, designs are listed WALTER L. SMITH, *University of North Carolina*.

Cox, D. R., "The mean and coefficient of variation of range in small samples from non-normal populations," *Biometrika*, 41 (1954), 469.

By examining special populations, a table is obtained for predicting approximately the mean and coefficient of variation of the range of random samples of any size up to five, drawn from a population of specified kurtosis The effect of non-normality on various statistical methods that use range is then considered [Author's Summary] WALTER L. SMITH, *University of North Carolina*.

Cox, D. R. and **Smith, W. L.**, "On the superposition of renewal processes," *Biometrika*, 41 (1954), 91–9

Suppose that there are a number of independent *sources* at each of which *events* occur from time to time The intervals between successive events at any one source are assumed to be independent random variables all with the same distribution, so that each source constitutes a renewal process Equilibrium behavior a long time from the beginning of the process for single and multiple sources is studied In particular, for a single source, expressions are obtained for (i) the delay function or distribution of time measured back to the immediately preceding event, (ii) the frequency function of the time interval between successive events and (iii) the probability distribution and variance of the number of events occurring in time intervals of a given length

Similar statistical properties are studied for the pooled output of N sources for both finite and infinite N The usefulness of the results for studying the underlying structure of a sequence of observed events resulting from a pooled output of N sources is illustrated with neuro-physiology data D G HORVITZ, *North Carolina State College*.

Cochran, William G., "The combination of estimates from different experiments," *Biometrics* 10 (1954), 101–27

This paper considers the problem of combining a number of estimates of a quantity μ The estimates may be called x_i, and with each x_i is given S_i^2 an unbiased estimate of the variance of x_i, based on n_i degrees of freedom.

The first problem to be considered is whether the x_i agree among themselves, if there is significant heterogeneity—i e , interaction, then the choice of a sensible method of combination requires careful thought.

Cases where the variance per observation is the same for all x_i are distinguished from cases where they are unequal. Rules are given for choosing from among weighted mean, semi-weighted mean, partially weighted mean, unweighted mean L E. MOSES, *Stanford University*

David, H. A., "The distribution of range in certain non-normal populations," *Biometrika*, 41 (1954), 463

The distribution of the sample range w_n in random samples of size n from certain special non-normal populations is considered Exact expressions for the expectation of w_n are obtained, and for the probability of w_n exceeding a given value On the basis of these and other results the author draws some general inferences concerning the effect of non-normality on the distribution of range WALTER L. SMITH, *University of North Carolina*.

David, H. A., Hartley, H. O., and Pearson, E. S., "The distribution of the ratio, in a single normal sample, of range to standard deviation," *Biometrika*, 41 (1954), 482.

This paper provides tables of upper and lower percentage points of the ratio, u, of the range of a sample of size n from a normal population to the standard deviation of that sample These tables are obtained by two methods of calculation The first uses known moments of the range and the standard deviation, of samples of size n, to calculate the moments of the ratio Pearson type curves are then fitted, and from these percentage points are calculated The second method is highly ingenious, and leads to exact upper percentage points. The ratio u is proposed as a test of homogeneity or normality of data, and numerical illustrations are given of its use in these connections WALTER L SMITH, *University of North Carolina.*

Dodge, Harold F., "Chain Sampling Inspection Plan," *Industrial Quality Control*, XI (January, 1955, 10-13)

Small sample sizes often require that the acceptance number equal zero Such attribute sampling plans have certain undesirable characteristics The author presents the following procedure as an alternative.

a) For each lot, select a sample of n units (n specimens) and test each unit for conformance to the specified requirement

b) Acceptance number of defects, $c=0$, except $c=1$ if no defects are found in the immediately preceding i samples of n ($i=1, 2, 3, \cdots$)

The equations for the probability of accepting a lot are derived and Operating Characteristic Curves are presented for $n=4$, 5, 6, and 10 and $i=1$, 2, 3, 4, 5 GERALD J. LIEBERMAN, *Stanford University*

Fucks, Wilhelm, "On nahordnung and fernordnung in samples of literary texts," *Biometrika*, 41 (1954), 116-32

This paper discusses various measures of the relationship of consecutive elements in a text (Nahordnung) and of distant elements (Fernordnung).

Nahordnung The text was split into consecutive groups of two words each and the number of syllables counted for each word. A matrix was constructed of the relative frequency of i syllables in one word and j syllables in the other word of the pair. Using this matrix, the mean number of syllables and variance for each word of the pair and the correlation between them were computed There are several misprints, and statements regarding the equality of means and the interpretation of the signs of the correlation are open to question Measures of skewness and entropy are also introduced Examples from several texts are included.

Fernordnung A matrix was constructed of the reciprocal of the average number of words (plus one) intervening between a word with i syllables and one with j syllables Various moments and other measures were also computed using this matrix. Another measure of the relation of distant elements was a correlation similar to the autocorrelation based on the number of syllables in consecutive words, except there was no correction for the mean If the number of syllables in consecutive words is designated as f_i, the correlation is presumably

$$q_l = \frac{\Sigma f_i f_{i+l}}{\sqrt{\Sigma f_i^2 \Sigma f_{i+l}^2}}, \quad i = 1, 2, \cdots, A - l,$$

where there are a total of A words The author uses a different notation for the summation, but the above seems to be his intent R L. ANDERSON, *North Carolina State College.*

Hammersley, J M., and Morton, K W., "The estimation of location and scale parameters from grouped data," *Biometrika*, 41 (1954) 296.

The distribution function $F(x)$ is unknown Successive experiments yield random samples from populations whose distribution functions are $F(\alpha x+\beta)$, the unknown parameters α and β varying from experiment to experiment. The observational data are grouped into not necessarily equal intervals It is desired to estimate the values assumed by α and β for each experiment, and to obtain as good an idea as possible of the true nature of $F(x)$ This paper provides a method of attacking these problems WALTER L SMITH, *University of North Carolina*

Hoel, Paul G., "A test for Markoff chains," *Biometrika*, 41 (1954), 430.

Bartlett has shown that certain frequency counts generated by Markoff chains are asymptotically normally distributed, and he was thus able to construct a likelihood ratio test for the goodness of fit of observational data. A feature of Bartlett's work is that the transition probabilities are sup-

posed to be known functions of certain parameters (finite in number) which can be estimated. The present paper constructs a goodness of fit test for the case in which the transition probabilities are completely unknown. The alternative hypotheses between which the test discriminates are (i) that a given observation depends only on the preceding r observations, i.e., the process is r-dependent, (ii) that the process is $r-1$-dependent. WALTER L SMITH, *University of North Carolina.*

J. H. Gaddum, "Bioassays and mathematics," *Pharmacological Reviews*, 5 (1953), 87–134.

This is a brief but surprisingly complete review of the various notions and methods peculiar to the statistics of bioassay. An excellent bibliography is included. L E. MOSES, *Stanford University.*

Fisher, Sir Ronald, "The analysis of variance with various binomial transformations," with discussion by M. S. Bartlett, F. J Anscombe, W G. Cochran, and Joseph Berkson, *Biometrics*, 10 (1954), 130–51.

In many experimental setups leading to an all-or-one response by the individual experimental unit, it is desirable to make inferences about some variate y, a function of the binomial parameter, P, associated with dose x. The probit, logit, angular transform, and others, are examples of such variates.

In each case estimation by maximum likelihood calls for an iterative procedure involving "working values" of the transform, and weighting coefficients, both of which change from cycle to cycle

The angular transformation and the square root transformation (for Poisson data), a limiting case of the former, have nearly constant variance independent of the parameter For both, the amount of information per observation is *exactly* constant. This property rather than approximate variance stabilization is offered as the actual justification for adopting the transformation. Approximate methods such as a final (non-iterative) analysis in terms of the empirical transforms are rejected as needlessly sacrificing an exact solution

In the discussion following the paper the point is often raised that most frequently there is present more variation than any of these models would imply The doubt thus cast on the precise applicability of the model reduces the desirability of trying to attain any "exact" solution. L. E. MOSES, *Stanford University.*

Green, J. R., "A confidence interval for variance components," *Annals of Mathematical Statistics*, 25 (1954), 671–85

The application of variance components in the interpretation of significance tests, the selection of efficient sampling designs and in other problems in genetics makes it worthwhile to have knowledge of the reliability of the estimates

Estimates for the variances of the variance component estimates are less reliable and less informative than confidence intervals.

The author stated his problem as follows Given two statistics μ_1 & μ_2 which are independently distributed as $\sigma_1^2\chi^2/r_1$ and $\sigma_2^2\chi^2/r_2$ with r_1 & r_2 degrees of freedom respectively. We want to find confidence limits for $\sigma_1^2-\sigma_2^2$(σ_1 & σ_2 are unknowns). Four solutions were obtained. The first was by a method similar to that used by Welch and Aspin on the problem of comparing two means and involves neglecting successively higher powers of the reciprocal of one of the degrees of freedom.

The second and third solutions were obtained by methods involving the neglect of successively increasing and decreasing powers, respectively, of a nuisance parameter The fourth solution, a more accurate one, was obtained by combining results in the second and third solutions. The accuracy of the different solutions was also discussed

To obtain a confidence limit in practice, the author discussed a tabulation which, as stated, is very laborious and has not been constructed A E SARHAN, *University of North Carolina.*

Kamat, A. R., "Distribution theory of two estimates for standard deviation based on second variate differences," *Biometrika*, 41 (1954), 1–11

This paper deals with the approximate distributions of the mean square and the mean successive second differences, δ_2^2 and d_2, respectively, where $(n-2)\delta_2^2=\sum(\Delta^2x_i)^2$ and $(n-2)d_2=\sum|\Delta^2x_i|$ for successive samples of n It is assumed that normal deviations with zero mean and variance σ^2 are superposed upon a quadratic trend, $\delta_2/\sqrt{6}$ and $d_2\sqrt{\pi}/12$ are then used as estimators of σ Values of the standard deviation, β_1 and β_2 for the distributions of δ_2^2 and d_2 are computed for $n=5, 7, 10$ (5) 30, 40, 50.

Exact distributions are given for $n=4$;

also the general characteristic equation of the matrix of the quadratic form for $(n-2)\delta_2^2$. R. L. ANDERSON, *North Carolina State College.*

Kendall, M. G., "Two problems in sets of measurements," *Biometrika*, 41 (1954), 560

If n random samples are drawn from a normal population of zero mean and unit variance, what is the variance of the sample of smallest absolute magnitude (i e., nearest the mean)? The author calls this the "Angel Problem" and provides the exact answer for samples of size 2, 3, 4, 5, and asymptotic formulas for higher n The results of the computations are tabulated alongside results of sampling experiments by W J. Youden. The author then considers a related "Demon Problem." WALTER L. SMITH, *University of North Carolina*

Mayne, Alan J., "Some further results in the theory of pedestrians and road traffic," *Biometrika*, 41 (1954), 375

Several statistical problems are tackled, concerning the passage of pedestrians across a road, on the assumptions that the time intervals between the arrival of pedestrians have one given distribution, that the time intervals between the arrivals of vehicles have another given distribution, and finally that all these time intervals are statistically independent In particular, the size of groups of pedestrians on an island between two lanes of traffic is discussed, and the increase in "efficiency" which can be obtained by the introduction of extra islands is calculated. WALTER L. SMITH, *University of North Carolina.*

Patankar, V. N., "The goodness of fit of frequency distributions obtained from stochastic processes," *Biometrika*, 41 (1954), 450

When the standard formula for the χ^2 statistic is used for grouped data certain independence conditions have to be fulfilled before we can resort to the usual tables to obtain a test of goodness of fit. When these independence conditions are not satisfied, as is the case with data arising from stochastic processes, we can either construct a more complicated test of goodness of fit, which takes account of the altered conditions, or we can study the distribution of the familiar χ^2 statistic when the usual independence conditions are not satisfied. Here the latter alternative is adopted and the effect of the serial dependence of observations on the classical χ^2 test is examined in some detail. On the basis of certain normality assumptions approximate distributions of χ^2 are obtained by the method of fitting moments. The results are applied to two important special processes. WALTER L SMITH, *University of North Carolina.*

Roy, S. N., "Some further results in simultaneous confidence interval estimation," *Annals of Mathematical Statistics*, 25 (1954), 752–61.

In continuation of the author's previous work along the same lines, confidence bounds are given on (i) the set of all characteristic roots of the dispersion matrix of a p-variate normal population (ii) the set of all characteristic roots of $\Sigma_1\Sigma_2^{-1}$ where Σ_1 & Σ_2 stand for the dispersion matrices of two p-variate normal populations (iii) all bilinear functions $a'\Sigma_{12}\Sigma_{22}^{-1}b$, where Σ_{12} stands for the covariance matrix between a p-set and a q-set and Σ_{22} for the dispersion matrix of a q-set in a $(p+q)$ variate normal population $(p\leq q)$ and a' is any arbitrary p-dimensional unit length row vector and b any arbitrary q-dimensional unit length column vector. In each case, the confidence bounds are given in terms of the observations and contain constants, with a joint confidence coefficient greater than or equal to a preassigned level The author considered some special interesting univariate and bivariate cases by putting $p=1$ in case (i) and $p=q=1$ in case (ii). He also showed that in case (i) "all characteristic roots of Σ" can be replaced by "all $a'\Sigma a$" and in case (ii) "all characteristic roots of $\Sigma_1\Sigma_2^{-1}$" can be replaced by "all $a'\Sigma_1 a/a'\Sigma_2 a$." A. E SARHAN, *University of North Carolina.*

Somerville, Paul N, "Some problems of optimum sampling," *Biometrika*, 41 (1954), 420.

This paper discusses the problem of deciding which of a given set of populations has the greater mean value, and considers the best balance between the opposing desiderata of economical sample size and of accuracy sufficient for the experimenter's ultimate purpose Certain general formulas are proposed to represent the cost of sampling and the loss incurred as a result of each of the various mistaken judgements which the experimenter could make. On the basis of this formulation, a theorem is proved which enables the experimenter to choose a sample size yielding the minimum expected loss. There is an application of the general results to the case where the sample means

are normally distributed. WALTER L SMITH, *University of North Carolina.*

Whittle, P., "On stationary processes in the plane," *Biometrika*, 41 (1954), 434.

The sampling theory of stationary processes in space is not completely analogous to the well-established theory of stationary time series, due to the fact that the variate of a time series is influenced only by past values, while for a spatial process the dependence extends in all directions This paper explores the consequences of this interesting fact and develops test and estimation procedures. The theory is applied to uniformity data for wheat and oranges WALTER L SMITH, *University of North Carolina*

Vaart, H R., van der, "Some remarks on the power function of Wilcoxon's test for the problem of two samples. I and II." *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen*, 53 (1950), 494–520, also *Indagationes Mathematicae*, 12 (1950), 146–172.

Vaart, H. R., van der, "An investigation on the power function of Wilcoxon's two sample test if the underlying distributions are not normal," *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen*, Series A, 56, No 5 (1953), 438–48, also *Indagationes Mathematicae*, 15, No 5 (1953) 438–48.

The problem considered by these two papers is that of testing that two populations, from which samples are drawn, are the same, against the alternative that they are the same except for a translation The alternative may be either one-sided or two-sided The tests considered are the one- and two-sided Wilcoxon (Mann-Whitney) tests The concepts of power function and unbiasedness are discussed in general, together with the relationship between the derivatives of the power function and bias.

In the two parts of the first paper the author derives general expressions for the power function of the Wilcoxon tests and their first two derivatives Numerical values of the first two derivatives at the null hypothesis of the Wilcoxon tests (for very small sample sizes) are compared with those of the two-sample t test when the distributions are normal These comparisons suggest that the Wilcoxon tests are only slightly less powerful than the t test for normal distributions, very small sample sizes, and small translations The extent to which these comparisons carry over to moderate sample sizes or to larger translations is not discussed It is shown that the two-sided Wilcoxon test may be biased

The second paper discusses the question of bias in more detail It is shown that the two-sided Wilcoxon test will be unbiased if the two samples are of the same size In the case of unequal sample sizes a sufficient, but not necessary, condition is found for the two-sided Wilcoxon test to be biased This condition is a two-fold one, one part being on the distribution functions, the other on the two sample sizes and the significance level of the test Two tables are given so that the sign of the bias can be determined for certain significance levels when the sample sizes are 10 and 5, or 8 and 7 JOHN P GILBERT, *University of Chicago*

BOOK REVIEWS

Statistical Theory of Extreme Values and Some Practical Applications. *E. J Gumbel* A Series of Lectures, National Bureau of Standards, Applied Mathematics Series 33 (Washington, D.C.: U S. Government Printing Office, 1954). Pp. viii, 51. $0 40 Paper. See review article by Bradford F. Kimball, on pages 517–528.

Introduction to Mathematical Statistics. Second edition. *Paul G. Hoel* New York: John Wiley and Sons, 1954 Pp vii, 331. $5 00.

ROBERT M. KOZELKA, *University of Nebraska*

HAVING been born and raised on the first edition of this book (studied from, assisted with, and taught from—at three different institutions), the reviewer approaches the second edition with a divided mind The new, more mathematical, chapters (2 and 3), which cover probability and the nature of statistical methods, are an important addition to the text. Furthermore, there is a stronger flavor of mathematics throughout, with alternative hypotheses and underlying sample spaces spelled out in each application. Another new chapter (10) on estimation and testing hypotheses, reënforces these ideas On the whole, the reviewer would feel much happier about his beginning statistical education if these features had been incorporated in the edition he studied from

The other side of the coin, however, was the feeling of having been mathematically short-changed upon finishing the book. The chapters designed to stress the applications of statistical ideas—the bulk of the book, and for the most part unchanged—have not made use of the mathematical approach much more than in the first edition. A new chapter (12), on design of experiments, left the same impression Furthermore, a number of minor faults in the first edition have not been corrected

One error in the preface (does anyone ever read the preface?) to this edition is the suggestion that those interested only in applications can omit Chapter 11 (Small sample distributions). Surely Chapter 10 is meant.

In a book of mathematics, it seems one should avoid empirical conclusions where possible In particular, in the normal approximation to the binomial distribution, Hoel still writes "Experience shows that the approximation is fairly good as long as $np > 5$ when $p \leqq \frac{1}{2}$, ." A more explicit statement would be that the mean is to be $np/\sqrt{npq}$ units from zero and hence $(np/\sqrt{npq}) > 3$ implies $np > 9 - 9p$. (Page 87).

His approach to some of the mathematical details, if not inexact, is at least confusing to this reader. On page 15 he states, " . . the random variable x is [now] just an ordinary variable of mathematics", and on page 39, " . . . the likelihood function [of observed values] gives the probability . . . ", which may lead an unsophisticated reader to wonder why this terminology was introduced in the first place. Further to bewilder the unsophisticated, he still does not give the normalizing constants of the F and t frequency functions. One could as justifiably omit $1/\sqrt{2\pi}$ in discussing the normal distribution, but one is hardly left with a text on mathematical statistics.

Probably the worst fault of all, from a pedagogical point of view, is the statement on page 11: "This result shows what is intuitively clear, ..." Is anything in any text intuitively clear to the average student? If it is, it is probably clear for an incorrect reason.

Lest one think that the book is nothing but a compounding of this type of thing, let the reviewer register some favorable impressions. The book dispenses a large amount of practical application with its theory, including many examples and exercises. For those of us unlucky (?) enough to have classes composed mainly of non-theoreticians, this is essential. Also, most of the objections of J. H. Curtiss, who reviewed the first edition in this Journal (June 1947), have been overcome. He stated, in part, "Any really serious adverse criticism would center about the treatment of regression, the omission of point estimation theory, and the handling of some of the proofs" The proofs remain simplified and non-rigorous, but the addition of Chapter 10 and the rather complete revision of Chapter 7 (Regression and correlation) should be satisfactory

The first edition of the book, was, in Curtiss' opinion, "In view of its general excellence, . the most important happening in the field of undergraduate statistical texts in recent years." This reviewer agrees, and concludes that the second edition, which he awaited eagerly, is even better. As a mathematician, however, he will continue to interpolate additional mathematics into lectures from this book, and await, eagerly, a third edition.

A Primer of Statistics for Political Scientists. *V. O Key, Jr.* New York: Thomas Y. Crowell Company, 1954. Pp x, 209 $2 50

GEORGE W. SNEDECOR, *Iowa State College*

THIS book is written for " . . the student who is unaware of the difference between a square root and a standard deviation . . . " (whatever that may mean). Five of the six chapters are devoted to "descriptive statistics"; frequency distributions, time series and interrelationships of time series, simple correlation and multiple relationships The sixth chapter is on "inferences from quantitative data " The illustrations are drawn almost entirely from the subject matter of political science.

In the first 5 chapters the author rigidly restricts himself to descriptions of populations. His readers will not be oversold on the usefulness of this approach because there are abundant warnings of limitations and pitfalls. The frequency distribution is the only device that gets unqualified approval. Measures of central tendency are of "extremely limited descriptive utility." Time series covering only 50 years are so short as to make it "most hazardous to develop a general cyclical theory " Of two series moving parallel, "Care is required in the interpretation of the relations shown to exist by even such simple methods." Concerning regression, "All the line does is to define the past direction; it does not probe into the future." In general, "statistical procedures will not do one's thinking for him," and "until one checks on the

ground he does not know that even the most plausible explanation drawn from the figures alone has a wisp of a foundation." In these chapters, while the author pays lip service to the effectiveness of the methods described, his constant warnings against fallacies and the paucity of constructive conclusions demonstrate the limited utility of his concept of descriptive statistics.

Teachers will be interested in the author's avoidance of numbers and calculations Graphs and verbal discussions constitute the bulk of the text. By this means the "ubiquity of the particular" is minimized I assume that this is partly because the audience is supposed to be unskilled in quantitative thinking. But I wonder if this has not been carried too far. Computation is a discipline which can ill be dispensed with Perhaps the most effective method of teaching lies between the two extremes

The last chapter is devoted to "Inferences from Quantitative Data." This is largely discursive, "without resort to a statistical formula." The t, F, and chi-square distributions are not mentioned In this discussion the author is not so happy as he was in his descriptions of populations Among 12 inferential statements which I listed only 2 were correct. This does not surprise me; if I were to undertake an exposition of elementary political science my blunders would doubtless be far more numerous The sad thing is that these 30 pages are largely wasted They might have been used to give adequate and usable methods for estimating and testing

In my opinion the author's distinction between descriptive and inferential statistics is irrelevant The only object of description is to facilitate inference The purpose of examining data is to estimate probabilities that enter into the making of decisions In practically all statistical investigations the inferences are extended beyond the data in hand, otherwise one would be wasting his time sorting dead bones The pertinent distinction is one that the author flirts with in chapter 6, the randomness or non-randomness of samples From random samples the inferences involve theoretically exact probabilities. Inferences from non-random samples imply faith in persistence or continuity, with probabilities that cannot easily be evaluated Random sampling is the appropriate subject matter of elementary statistics The graver problems of non-random sampling should be left to professionals in the specialized disciplines involved

I respect the self-control that enabled the author to refrain from any sample-to-population inference in his first five chapters but I think this restraint removed these chapters from the domain of statistics Mere arithmetical and graphical descriptions of data have no more than historical value. From the reviewer's standpoint, the chief interest in this text is not the statistics, which is of doubtful value, but the sage, humorous, and critical attitude of a political scientist toward the fascinating problems which he presents

P.S. I am sorry I couldn't manage a judicious use of the word *parameters* because the author says that this is "a term which, judiciously used, creates an appearance of erudition "

Design and Analysis of Industrial Experiments. Edited by *Owen L Davies* Authors· *G. E P Box, L R. Connor, W R. Cousins, O. L. Davies, F. R. Himsworth,* and *G P Sillitto* New York· Hafner Publishing Company, 1954 Pp xiii, 636. $10 00

HAROLD A. FREEMAN, *Massachusetts Institute of Technology*

FOR years we waited for a good elementary textbook in mathematical statistics, along came, in rapid succession, Hoel, Weatherburn, and Mood. For years the field of sampling surveys had practically no book literature; now we have Deming, Hansen-Hurwitz-Madow, Cochran, Sukhatme, and Yates. The book situation was equally barren in experimental statistics, particularly as applied to batch (chemical) processes Now, following Brownlee and Youden, we have Gore and, within weeks of each other, two comprehensive volumes, one by Bennett and Franklin (to be reviewed in this *Journal*) and the volume discussed here

Davies' is a statistical handbook for experimenters and a very fine handbook it is. The applications lie almost without exception in the general area of chemical processes but one does not need much experience or imagination to see that problems from many other areas can be readily described by the models given in the present volume These models arise from the practical problems considered in the text—many of them encountered by the authors in their own researches—and are naturally specialized to these problems. But the reader with modest training in mathematical statistics will often be able to modify them to his needs

The technique of this volume is literary, with formulas and examples There is some proving, most of it confined to appendices and particularly careful attention is paid to the assumptions underlying the models. The exposition is clear and accurate, and many topics of importance to experimenters—and often neglected in the textbooks—are found here These include, for examples, lack of independence of errors, self-containment in experiments, sample sizes as functions of sampling and test costs, estimation of variances rather than the more usual formal analysis of variance, effect of departure from normality, missing values, Latin cubes, combining interaction mean squares.

The body of the book is naturally given to experimental designs of the Fisherian sort, all centering on variance analysis, for little else seems to have made any mark in experimental statistics The topics include simple comparisons, randomized blocks, Latin squares, incomplete blocks, factorial designs, confounding, and fractional factorial designs. Much of the content can probably be found elsewhere but it is good to find it in one place in such excellent form There is also a fine chapter on sequential tests, including the details of (Barnard's) sequential *t.* The most interesting chapter, to this reviewer, was 11, on the Box-Wilson technique—designs for multi-factor experiments which aim at the optimal value of a characteristic, for example, the levels of the ingredients which lead to the maximum strength of an alloy. So far as I know, this is the first detailed account in book form of this

most interesting work The numerical details are involved but as the method contributes powerfully to the solution of a problem often encountered by experimenters, it has a certain future and many experimenters will want to master it. They will find a thorough account of it here.

This will be a book well worth keeping up to date, and it is the authors' intention to do that.

Statistics for Technologists. *C G Paradine* and *B H P Rivett* New York: D Van Nostrand Company, Inc, 1954 Pp vii, 288 $6 75

GERALD J LIEBERMAN, *Stanford University*

THE widespread recognition of the applicability of statistics to problems in engineering and the physical sciences has demonstrated the need for educating the technologists in the field of statistics. As a result, numerous books in the field will be published, *Statistics for Technologists* being one of the first few.

The content of the book is typical of standard texts for a first course in statistics. The preliminary chapters deal with frequency distributions, probability, and probability distributions Tests of significance are presented for various hypotheses about parameters of these probability distributions. The book also contains brief chapters on quality control and sampling inspection. The theory of errors and the method of least squares are presented together with illustrations of their use The remaining chapters deal with correlation, analysis of variance, probit analysis, and the principle of maximum likelihood

The authors are to be commended on their point of view regarding the level of mathematics used throughout the text Their attitude is best expressed in their own words

"During the last decade or two there has been a considerable raising of the standard of mathematics required of candidates for degrees or higher national certificates in science and engineering The students are used to a mathematical presentation of the theory of their own subject In the past many books on statistical methods have given formulas and rules of procedure without going into the underlying mathematical theory In this book we try to steer a middle course, giving the mathematical derivation of most of the results required, with an introduction to their applications, in a sequence which it is hoped will appear logical and carry conviction Although there can be no objection to taking the mathematics for granted if the operation of statistical methods is properly understood, the fact that statistical theory is still advancing makes it the more desirable that the basis of standard procedures should be revealed The student who has followed the arguments used in establishing a sampling distribution, for instance, is more likely to be aware of assumptions made in applying it and better equipped to read further in the subject."

The book contains many interesting examples and problems Some of these

are taken from examination papers of the Senate of London University, the Royal Statistical Society, and the Association of Incorporated Statisticians.

Statistics for Technologists is supposed to be an introduction for students, research workers, and engineers to the principles of statistical methods and theory. These people, however, will find it lacking in many respects There is almost no discussion of how to design an experiment. Tests of significance are treated only with respect to the error of type 1. The error of type 2 is never mentioned. This is paradoxical since in the chapter on quality control, the authors do devote space to determining the sample size necessary to detect a shift in the process average with a given probability.

The chapter on sampling inspection is meager, and contains very little information about such important topics as sampling inspection by variables. This may not be very serious except that the references to other sources are very poor No mention is made of such works as *Sampling Inspection by Variables* by A. H. Bowker and H P Goode and *Statistical Quality Control* by E. L. Grant This is not only characteristic of this chapter but of almost all the chapters It appears that the authors are not aware of many of these references They state in the preface, for example that "no originality can be claimed for the methods adopted, but it is thought that the use of the χ^2 table for the determination of a single sampling inspection scheme (page 144) may not have been previously noted" This result has appeared at least twice in the literature[1]

The book includes the topic of point estimation but hardly discusses the very important concept of confidence interval estimation To the scientist this is an absolute must. Naturally this precludes the introduction of the new work on comparing means in the analysis of variance presented recently in the literature by Tukey and Scheffé

After reading the book, one gets the impression that this is a book which will help many a scientist, even though it lacks many of the qualities that are necessary to make it an "extra" good book It presents the standard topics that are included in every text. However, it lacks many topics which are included in most texts, topics which are necessary to answer important questions that the scientist will ask

---

Quality Control: A Manual of Quality Control Procedure Based upon Scientific Principles and Simplified for Practical Application in Various Types of Manufacturing Plants. Second Edition. *Norbert L Enrick* New York Industrial Press, 1954 Pp viii, 181 $4 00 (Brighton, England Machinery Publishing Co, Ltd)

GEOFFREY GREGORY, *Stanford University*

IN THIS, the second edition of the book, Enrick has added five chapters to the original work, with the idea of introducing the reader to some more specialized techniques of the subject Essentially Part I is the same as the

[1] Paul Peach and Sebastian B Littauer, "A note on sampling inspection," *Annals of Mathematical Statistics*, 17 (1946), 81–4, Joseph Cameron "Tables for constructing and for computing the operating characteristics of single sampling plans," *Industrial Quality Control*, 9 (1952), 37–9

first edition of the book, which has been reviewed by John H. Curtiss in this *Journal*.[1] Additional explanatory passages have been inserted, however, and indications have been made of further developments of topics in the second part of the book. In particular, the author elaborates on the sampling risks of his sequential sampling plans, but it is perhaps a pity that some light could not have been shed over his use of the term "allowable per cent defective," a major consideration in selecting a sampling plan. This additional material is well worth studying, however, if only as an indication of the care which should be exercised when a particular plan is selected. Also, considering the wide use made at present of the Military Standards and the Dodge-Romig tables, it would have been of great assistance to the aspiring quality control engineer if at least some introduction to the terms used in selecting these plans could have been made

Continuing in his pleasantly easy style, in Part II of the book Enrick launches out into some of the more sophisticated techniques of the subject His decision in the foreword to the first edition to design the book for practical men in inspection, and so eliminate any hint of "higher mathematics" here leads him into some difficulty. Such examples as testing for normality or the comparison of variations are described in principle but left without any precise procedure for making a decision A similar observation may be made of the treatment of the analysis of variance It would be impossible, of course, to describe the full implications of the analysis of variance in a book of this type He does succeed, however, in giving some indication of the possibilities of these tests, and no doubt the inspector, new to these tecnhiques, will find it a very useful introduction into their scope As the author points out, some further instruction or assistance is necessary before application is made

The remaining added material is devoted mainly to a description of the conventional Shewhart control charts which had been curiously omitted from the first edition A question might be raised about the significance level of different charts, since we are apparently using 2σ limits for the means charts and 3σ limits for the range charts The standard procedure is used for the treatment of the attributes control charts

In its present form the book sets down in a simple manner the basic techniques of statistical quality control It will be of interest to the non-technical man who wishes to obtain some insight into the possibilities and procedures of the subject Throughout the book the theory is illustrated with examples of admirable clarity. To understand the techniques well enough to apply them, however, rather more than this is needed Quality control engineering is a dangerous occupation for the amateur As a textbook, then, this treatise can be considered as an introduction only to the recognized standard texts on the subject As such it is well written and to the man who entertains doubts about the usefulness of the subject it has the virtue of being written by a man with considerable practical experience

[1] Volume 44 (1949), 139–141

Definitions and Symbols for Control Charts. New York. American Society for Quality Control, 1951. Pp 8 $0.50

EDWIN G. OLDS, *Carnegie Institute of Technology*

THE present document represents the work of a Committee of the American Society for Quality Control. It has been approved as a standard by that organization.

Arranged in matrix form, for ready reference, this standard gives the terms, symbols and definitions for use in connection with control charts. The first section has fifteen general terms relating to control charts, the second section has twelve terms relating to control charts for variables, and the third section has terms relating to control charts for attributes

The Committee is to be commended for bringing together many of the principal ingredients of control-chart communication. With some general agreement on meaning and symbolism, mutual understanding should be improved and work in the field should benefit

Writing definitions is a tricky business because words must be defined in terms of other words which then may need further definition It is difficult to decide where to start and when to stop In the present instance, it is surprising that the committee stopped before defining *control*, or *quality control*, or *statistical quality control*. Perhaps some definition of control is implied in the statement, "Assignable causes must be identified and removed to attain statistical control " However, a much more direct attack on the difficult problem seems to be required. Other surprising omissions are *chance cause, statistic, parameter, modified control limits, variance, expected value,* and *random sample.*

Most of the definitions given differ little from those to be found in quality control publications In many cases the definition includes an example or a statement of use Some readers may wish to object to the definition of "average" as the arithmetic mean and may find other definitions which could be made more precise In general, however, the definitions given can be expected to convey correct notions, with the possible exception of the dual formulas for the standard deviation of p, pn, u, or c

When proposing a standard set of symbols for use in a particular field there seem to be at least three questions which might be raised.

(1) Are the proposed symbols those which have already been widely used?

(2) Will adoption of the proposed symbolism promote better communication with workers in neighboring fields?

(3) Is the pattern of symbolism such as to allow easy extension as the work in the field deepens and broadens?

In answer to the first question, it might be noted that the present standard retains, with little change, the set of symbols used in the A S T.M. manuals, *On Presentation of Data*, and *On Quality Control of Materials*. The prime notation is "used to designate the true or objective value " This is the same convention used in the American Standards Association's publication *Control Chart Method for Controlling Quality During Production*, one of the most

widely-used guides for the application of the control-chart procedure. Sigma is used, as in the above publication, to denote the sample standard deviation. However, a footnote kindly authorizes two departures from standard: (a) the use of the single prime notation may be restricted to universe values and a double-prime notion used for standard values and (b) Fisher's *s* may be used for control charts for standard deviation.

While the suggested symbolism bears the stamp of approval for past usage, it has created, and promises to continue to create, problems with regard to both intercommunication and growth Many students of statistical quality control believe that there is an opportunity to apply a variety of statistical methods in the improvement of the manufacturing operation. When they move out of the area of simple control charts they are faced with the need to learn a new notation. This reviewer has observed the anguished struggles with this difficulty and he wishes that it could be avoided

Probably the committee gave sober consideration to the difficulty noted above but it is not clear that its decision was the best one. Perhaps it would have been better to wait and hope that the statistical world would meet in a common symbolism which would well serve the needs of all specialists. The formal adoption of the proposed symbolism as a standard sets up a barrier to this much needed development which may be hard to hurdle However, any adverse criticism should be tempered with the recognition that progress toward a unified symbolism in the field of statistics has been painfully slow. Possibly the ASQC action will revive interest in the project and will speed developments

---

Statistical Astronomy. *Robert J. Trumpler* (Professor of Astronomy, Emeritus, University of California) and *Harold F. Weaver* (Associate Professor of Astronomy, University of California) Berkeley and Los Angeles University of California Press, 1953 Pp xiv, 644 $7 50

W EDWARDS DEMING, *New York University*

THE aim of the book (not stated) is to explain in the language of the astronomer, and to the astronomer, some of the methods of statistics and their uses in astronomy Looking back at history, one might say that the aim is to acquaint astronomers with their own methods, because astronomy, like other sciences, is statistical in its quantitative problems Moreover, it was astronomers like Gauss, Legendre, Czuber, and Helmert who made great advances in statistical theory as a result of recognizing the nature of their problems In the last three decades, however, statistical theorists have pushed ahead, so that now it is the astronomers who find themselves in need of a compendium of recent statistical theory—a field in which they were at one time preeminent.

The statistical problems in astronomy deal with the estimation of the parameters of orbits, with the ratios of weights, with velocities, periodicities, chemical compositions, calibration of instruments, the space-distribution of

stars, measurement of t and the ities of light of various wave lengths, measurement of other spectr s are cteristics, problems of curve fitting, interrelationships between th ractical characteristics of stars and of galaxies. In addition, there are the statistical problems that confront the experimental scientist everywhere the discovery and removal of systematic errors and biases, and the design of experiments so as to extract the greatest amount of information from whatever observations it may be possible to take.

All the various statistical techniques are useful in astronomy—distributions, theory of sampling, testing of hypotheses, analysis of variance, design of experiment, tests for randomness, statistical theory of prediction, of functional relationships.

It is now known to many physical scientists that the theory of statistical design and of inference are a vital part of precise experimental work Even the greatest experts in the most refined physical measurements find that statistical theory helps them to discover and remove systematic errors and biases, and helps them to plan an experiment with the correct amount of replication

Part I in 8 chapters deals with statistical theory on an intermediate level, all the while with examples in astronomy, and in the vocabulary of the astronomer. Part II in 3 chapters deals with the statistical description of the galactic system Part III in 6 chapters deals with stellar motions in the vicinity of the sun Part IV in 3 chapters deals with luminosity. Part V in 5 chapters deals with the space-distribution of stars Part VI in 3 chapters deals with galactic rotation

A special feature of the book, to the statistician, is Chapter 8 in Part I on the testing of hypotheses, contributed by Elizabeth Scott In the other chapters of Part I there is considerable emphasis on the transformation o variables and to relations between frequency functions The mature statis cian and the beginner will both find much of interest here, as well as in whole book

The authors may give the impression in the beginning lines of the ntı duction that a statistical investigation is concerned with the quant tati description of a group of objects or individuals, or that it is to gain ir iform tion on the distributions and interrelations of certain attributes th at cha acterize the individuals of the group. They can not be blamed a p rominer text-book in statistics written by statisticians uses almost these sa ime word to describe the statistical method. Fortunately, by the time the astronome has read to the middle of the first page of the introduction he ma y perceive that the foregoing description of a statistical investigation is hap nly not so after all; that the real aim in modern statistical theory is to *discover the causes* of the distributions and of the interrelationships Elizabeth Scott in her chapter is clear "we should like to make the correct decision as often as possible", and statistical theory helps the astronomer or anyone else to do this

Page 237 will bring smiles to a statistician who has engaged in studies of

human populations. We read there that the introdu ment of information from every member of a human population is co not gr tively simple. All that he needs is a suitable questionnaire (easily de topid, no doubt), and there is apparently no trouble in finding all the peo ike and in persuading them to answer. Then begins the work of distributions and interrelations. In contrast, the astronomer has difficulty in procuring the information from his stars! All of which reminds me of the lawyer who remarked recently that a census of employment and of unemployment should be the simplest thing in the world! Everybody else's job *is* simple.

The book and 1953, its date, will doubtless be known, and justly, as a great achievement in astronomy and in the other physical sciences. It is recommended to the statistician who wishes to learn his subject from the standpoint of the scientist and to see statistical theory applied to the universe of stars and of galaxies, from the earth to the outermost reaches of the astronomer's vision, 100,000,000 or more light-years away.

Table of Binomial Coefficients. Prepared for the Mathematical Tables Committees of the British Association and the Royal Society, under the editorship of *J. C. P. Miller.* Royal Society Mathematical Tables, Volume 3 Cambridge. Published for the Royal Society at the University Press, 1954 Pp. viii, 162. $6 50

THIS table shows the number of combinations of n things r at a time for all values of r and n such that $r \leqq n/2 \leqq 100$. This in effect covers all values through $n=200$ For each n, the maximum value is (or values are) printed in **boldface**. The tabulated values are exact, including as many as 59 significant figures. In addition, the table covers values of $r \leqq 12$ for $n \leqq 500$, $r \leqq 11$ for $n \leqq 1000$, $r \leqq 5$ for $n \leqq 2000$, and $r \leqq 3$ for $n \leqq 5000$. Reference is given to a table published in 1762 covering $r \leqq 2$ for $n \leqq 20{,}000$.

W. A. W.

Introductory College Mathematics. *Adele Leonhardy* New York· John Wiley and Sons Pp ix, 459. $4 90

D A. DARLING, *University of Michigan*

IN HER own words the author has designed this book "primarily to meet the needs of the college student who does not plan to specialize in mathematics or the related sciences " It is therefore potentially of interest to teachers of statistics. The prerequisites are a minimum of one year of high-school algebra and one year of plane geometry. The main emphasis is "on what and why and not merely how", and generally speaking Miss Leonhardy strives for a broad and mature understanding rather than for the development of manipulative skills.

To carry out this program the author treats the historical development of mathematics and stresses often the arbitrary nature of its postulates. Thus there is a discussion of non-Euclidean geometries, numbers expressed in non-denary systems, non-commutative algebras, etc. There is a treatment

of elementary set algebra and the related syllogistic system. Operations with the class of positive integers are treated axiomatically, as are their extensions to negative integers and "fractions." The first three chapters deal with algebra through roots and powers. Chapters 4 and 5 are on measurement, computation, logarithms, ratio, percentage, index numbers. Chapters 6, 7, and 8 treat variables, functions, functional relationship, curve plotting, simple polynomials, roots, nomographs, variation, limits, derivatives (to the power function), antiderivatives (of polynomials). Chapters 9 and 10 treat the exponential, logarithmic, and periodic functions

Of interest to readers of this journal is the penultimate chapter, 11, entitled "Simple Statistical Methods" The author defines "mathematical (or a priori) probability" by the "equally likely" decomposition and distinguishes this from "statistical (or a posteriori) probability" based on empirical frequencies. She says of the latter definition that "a large enough number of cases must be observed so that the conclusions can be sound" There is a description of frequency charts, mean, median, mode with a number of illustrations followed by a formula called the "normal frequency curve". There is next a definition of the standard deviation and a paragraph or so discussing "correlation"

The book closes with a chapter sketching the development of mathematics and its possible future.

Certainly any teacher would give the author his best wishes for the success of this book, and would like to believe that the beginning student could gain "a rich background in mathematics" from a course based on it. There is a great deal of chatty discussion throughout and informal, lengthy motivation. There are numerous problems, many of the "thought" type· "What is a proposition?", "What are the materials of mathematics?", "What are the natural numbers", "Discuss the law of large numbers in probability", "Define the indefinite integral of a function in terms of symbols." All in all, there is a great deal of reading to be done by the student and at the end of each chapter there is a supplementary bibliography which in many cases, in the reviewer's opinion, is unrealistic to suggest to a recent high school graduate. For example, at the end of the first chapter there are references to Carnap's "Foundations of Logic and Mathematics" and Hilbert's "Principles of Mathematical Logic".

It is not difficult to find many technical faults, but it is perhaps not germane to do so with a book of this type One might remark that the author really does not have a postulational treatment of the integers, etc , at all despite her lengthy and meticulous attempts. It also seems unfortunate that she has had to refer in various places for simple proofs to texts in college algebra and has developed insufficient machinery, for example, even to state the binomial theorem There are also the usual fuzzy definitions of limits, derivatives, etc.

As a preparation for later statistical work by the student, the reviewer feels that this book is a step in the right direction in that set theory, prob-

ability, and the basic statistical ideas are introduced to the student early, but it is somewhat doubtful that he would not gain as much or more from the standard courses, supplemented by these topics. The reviewer personally would like to try a course based on a book like this to resolve the doubt.

Career perspectives in a bureaucratic setting *Dwaine Marvick.* Michigan Governmental Studies, No 27 Ann Arbor: University of Michigan Press, 1954 Pp viii, 150. $2 25

DAEL WOLFLE, *American Association for the Advancement of Science*

Two hundred and four employees of the Office of Naval Research were studied by means of a two- or three-hour interview and a questionnaire in which each employee supplied information concerning his personal history, education, career plans, and values The employees were classified in groups of almost equal size as follows "institutionalists"—civil service or military persons who looked forward to careers in government service; "specialists" —scientists, engineers, accountants, lawyers, who looked forward to careers in their specialties rather than in a particular place of employment, and "hybrids"—who were both government service and specialty persons and who looked forward to careers combining both interests.

Various predictions concerning the three groups were made, e g , that the specialists would be more likely than the institutionalists to consider opportunity to learn and opportunity to do original work as very important factors in their jobs and that the institutionalists would be more likely than the specialists to consider security, salary, and organizational prestige as very important Most of the predictions were borne out by the interview and questionnaire data

Results of the study are given in 33 tables, 12 figures, and too many words. The monograph ends with a brief discussion of implications for management As an example since specialists are not very firmly attached to a particular place of employment, management could appropriately "bend its efforts toward inculcating a greater attachment to the institutional place and toward stressing the relative advantages of this agency."

Local Community Fact Book for Chicago, 1950. Edited by *Philip M Hauser* and *Evelyn M. Kitagawa* Chicago· Chicago Community Inventory, University of Chicago, 1953 Pp xx, 310 $25 00

SANFORD M. DORNBUSCH, *University of Washington*

After the decennial census of 1930, statistics for the 75 community areas of Chicago were presented for the first time in a single volume. The success of the initial venture led to its repetition using data from the 1940 census, but the second *Fact Book* did not appear until 1949 The editors of this third member of the series deserve credit for its speedy publication.

Their haste has not led to any decline in quality, for the data are complete and well-organized.

In addition to published census materials, the *Fact Book* incorporates the results of special tabulations from census tract summary cards and data from local agencies. This volume presents information on many characteristics not covered in preceding editions; namely, class of worker, major industry group, family income, migration status, tuberculosis rate, public assistance rate, number of persons in the dwelling unit, type of fuel, presence of television, mortgage status of owner-occupied homes, number of married couples not living in their own household, and gross land area. A final noteworthy inclusion is material on retail stores and retail trade taken from the 1948 Census of Business.

A minor deficiency of this new version is the deletion of the sex breakdown of material on individual income. Although the new information on family income is most welcome, it does not permit examination of the changing pattern of income distribution within the female component of the labor force.

The Negro population of Chicago increased 77.2 per cent from 1940 to 1950, and the changed organization of the *Fact Book* reflects this rapid rate of growth. Color breakdowns of data have been added for the following kinds of material: age, marital status, school years completed, type of household, employment status, major occupation group, characteristics of dwelling units, births, and deaths.

A major improvement is the addition of a history of each community area. The recent developments which are noted illustrate the increasing heterogeneity within community areas. This is the price that must be paid to preserve comparability for trend analysis. This reviewer does not agree with the editors' view of most of the community areas as containing persons who are aware of common community interests. The conception of the community area as a natural area would call for radical redefinition of boundaries to allow for recent migration.

Labor Mobility in Six Cities. *Gladys L. Palmer* New York: Social Science Research Council, 1954 Pp 177 $2 75

David R Roberts, *Carnegie Institute of Technology*

Labor *Mobility in Six Cities* is a landmark in the study of mobility. Its significance can be fully perceived only in historical perspective. Very little was known about the extent, the incidence, or the pattern of worker mobility prior to the 1930's. Economists who might have initiated research were uninterested because one of the simplifying assumptions of the accepted body of theory was the perfect mobility of all input factors including labor. The existence of immobility was recognized but it was treated as a friction which would modify theoretic results but would not invalidate them. It

was not deemed necessary to engage in empirical investigations of the phenomenon. Psychology and sociology, being concerned with individual and group behavior, were in principle interested in job mobility, but psychologists and sociologists were preoccupied with research in other parts of their fields. Practical people, such as businessmen and government officials, were not concerned with mobility because it did not pose problems for them. During the 1920's some business organizations became interested in reducing labor turnover as a means of cutting costs, but because of the narrowness of their objective they did not learn much about mobility per se.

During the 1930's the practical problem of administering relief highlighted both the need for and the lack of knowledge about labor mobility. For example, to what extent can people be expected to leave depressed industries, occupations, and areas? Which groups move and which stay? etc. During this period and the early 1940's, a number of mobility studies were made; they were all small and sought to illuminate the labor market behavior of particular groups of workers. One, for example, studied the mobility of weavers in three textile centers, another the mobility of former employees of a large textile mill which had failed; still another, the reemployment of Philadelphia hosiery workers after the 1933–34 shut downs Like the early experiments in other areas of scientific inquiry these studies provided a little factual information but primarily they suggested hypotheses, possible explanations of who does and who does not change jobs, the areas within which changes occur, etc Substantively, they were not of much general use because their coverage was too narrow a base for generalization and their findings were frequently inconclusive or inconsistent with those of other studies having as good claims to acceptance

This was the state of knowledge prior to the present study. It is a large empirical undertaking conducted under the leadership of Gladys L. Palmer for the Labor Market Research Committee of the Social Science Research Council, drawing upon the facilities of seven university research centers and the Bureau of the Census The basic data were provided by a statistically designed sample survey of the populations of six northern, industrial cities: Philadelphia, New Haven, Chicago, St Paul, San Francisco, and Los Angeles The field work was done by the Census Bureau and the analysis by groups at the Universities of California, Chicago, Minnesota, Pennsylvania, Yale University, and Massachusetts Institute of Technology Work histories were obtained covering the decade of the 1940's for individuals 25 or more years of age who had worked for a month or more in the year 1950. The findings are based upon the analysis of those work histories. They are far too numerous and technical for detailed presentation. The major ones follow:

1. Mobility is not characteristic of all members of a labor force but is concentrated within certain parts.
2. Differences in the incidence of mobility among different groups of workers and the kinds of job shifts made follow a similar pattern in

different cities, regardless of whether a city's degree of mobility is relatively high or low.
3. There are differentials in the incidence of mobility at various levels of skill, but even highly stable occupational groups have mobile segments.
4 A labor force adapts more readily to changes in the industrial demand for labor than to changes in the occupational structure
5. Persistent intercity contrasts suggest the existence of area differentials in mobility.
6. Expanding employment in a city attracts workers from other areas, and migrants are relatively flexible in adjusting to labor market changes.
7 Workers who are experienced in certain occupations can transfer their skills to certain others, but there is a limit to the amount of interchange between levels of skill.
8 When employment is high, voluntary job changes outnumber involuntary changes and tend to reflect an improvement in economic position and in the knowledge and skills of workers.

There is a wealth of statistical breakdowns resulting from the cross-classification of the sample by city, age, sex, race, marital status, occupation, industry, employer shifts, industry shifts, occupational shifts, migration, etc.

Rather than attempt a recapitulation, it may be more fruitful to consider the two following questions (1) To what extent did the study succeed in determining the extent, incidence, and patterns of mobility within the area examined? (2) How far can its findings be generalized? With respect to the first question, the results are mixed. Findings about the extent and patterns of mobility are conclusive and constitute an important contribution to knowledge Never before has there been such comprehensive information, even for a single area, on the amount of job movement of its population and on the character of the movement, i e, how it breaks down into employer shifts, industry shifts, occupation shifts, geographic shifts and combinations of these.

The findings about incidence are less satisfactory. They establish that a large portion of the job shifts are made by a small fraction of the workers—a finding which had been foreshadowed by earlier studies It establishes that mobility varies inversely with age and it suggests strongly that the people of higher socioeconomic rank are, in general, less mobile than those of lower rank. However, the attempt to distinguish the mobile people further, in terms of sex, race, industry, etc., yielded disappointing results The inconclusiveness of the findings shows up in the fact that significant differences in mobility rates do not appear when the data are classified according to sex, according to race, etc. The design of the study tends to obscure such differences to some degree. The people in the sample were classified according to their 1950 characteristics and their work histories for the whole preceding decade were then imputed to those characteristics For example, if a man married for the first time in 1950 his job history for the preceding decade

would be associated with the status "married," even though nine-tenths of it pertained to the status "single." Another factor of importance is the size of the sample used. In order to isolate the effect of any single characteristic, e.g., sex, upon mobility it is necessary to compare the mobility rates of groups of people who are similar in all significant respects except sex, e.g., age, occupation, industry, etc. Such extensive cross-classification is impossible within the limits of statistical reliability unless the sample is much larger than that employed in the six-city study. It was necessary, carrying forward the foregoing example, to compare mobility rates for men and women without providing uniformity in respect to age, occupation, etc. To the extent that differences in the latter characteristics influence mobility the true effect of sex does not show up. The omission from the sample, probably for financial reasons, of people who lived in the six cities at some time during the decade but moved away before the survey date, biases the results and may be responsible for some of the inconclusiveness about incidence, though this factor is less important than those mentioned above.

The question of generalizing the six-city findings is complex. In principle, generalization is valid only for areas which are similar in essential respects to those used in the original study. A difficulty is that in the present state of knowledge it is not at all clear which aspects of an area are essential in respect to the mobility of its inhabitants. Only additional studies based on areas which differ from the six cities can supply that information. In the meanwhile generalization must be based on judgment. Regardless of the generality of the six-city results, the study has provided a set of benchmarks and some notions about mobility which will be a valuable frame of reference for future workers. A less tangible, though potentially significant, contribution is the guidance which the study provides for future research in the form of now-recognized hypotheses which were not apparent until illuminated by the evidence amassed.

Without question, this if the most important piece of work in the field and one whose significance must be measured both in terms of its substantive contribution to knowledge and the influence which it will exert on future research.

Industrial Pensions. *C. L. Dearing.* Washington: The Brookings Institution, 1954. Pp. x, 310. $3.75.

H. W. Steinhaus, *The Equitable Life Assurance Society*

When the United States Supreme Court, on April 25, 1949, upheld the Inland Steel decision which made pensions subject to collective bargaining under the Taft-Hartley Act, economists familiar with pension fund accumulations became concerned with the possible consequences if organized labor were to succeed in obtaining substantial pension rights which in turn would lead to tremendous pension reserves. The Brookings Institution began a study of this formidable problem early in 1950 and published its con-

clusions in May 1954, some 4 years later. It was an exploratory undertaking in a new field, made particularly difficult by the need to evaluate and project numerous factors which enter into the creation of private retirement security on a national scale.

These factors involve (1) estimates of the future number of aged, their longevity and their private resources, which would influence the trend of pension demands; (2) estimates of the number of people in the labor force, their earning power and productivity, which would determine the capacity to absorb the cost; (3) the relation of savings for retirement to savings generally, and the amount of such savings in relation to investment demands; and (4) the evaluation of the trend in pension plans, as to inclusiveness, degree of vesting, method of financing, ability of unions to promote, ability of employers to resist, and government influence through taxation or publicly administered schemes.

Unfortunately, there is no point in discussing the results of any of these estimates, since most of the basic premises used for projections have already become obsolete. There have been fundamental changes in the Social Security Act and the Revenue Code The savings figures of 1949 and prior years, in total as well as by income brackets, which were used for projections have proved invalid in the light of more recent experience. Such newer developments as investments in equities, portable pensions, and the change in union emphasis on the tie-in between private and social pensions, were not evaluated Unwittingly perhaps, but forcefully, the book demonstrates the futility of attempting projections in an economy as dynamic as ours

Nevertheless, the student of the subject will find much of interest in a review of the techniques developed for the various estimates Economists may benefit from the painstaking search into methods to obtain a rough estimate of the future of industrial expansion which in turn influences both the supply of funds and the investment potentials. Chapter VII, for instance, estimates the annual flow of money into pension funds, by rate, volume, and characteristics, by utilizing data supplied by 297 corporations on a special questionnaire. Because of great variations, each major industry was analyzed separately, and it was concluded that increases in retirement savings will not be fully offset by decreases in other forms of savings.

In Chapter VIII a projection was made of national income changes and of personal savings particularly, and then a detailed analysis was undertaken of future investment outlets in the form of federal, state, and local government debt increases, mortgage potentials, and corporate security issues. This involves, in the public sector, construction of buildings for public education, highways, and urban redevelopments; in the private sector, residences, plant expansion, and new industries It was concluded that some portions of of the total supply of new retirement savings might be unable to find ready employment.

Chapter IX appraises the basic conflicts in policy and objectives that have characterized the pension development and considers how these conflicts may be resolved. The final chapter, X, attempts to resolve the vital issue of

allocating responsibility for retirement financing. In spite of all shortcomings, these four chapters contain challenging thoughts and open up new avenues for appraising the economic effects of retirement security established through private means. Above all the study convincingly demonstrates that retirement security on a national scale affects every phase of our economic life.

The Share of Financial Intermediaries in National Wealth and National Assets, 1900–1949. Occasional Paper #42, Studies in Capital Formation and Financing. *Raymond W Goldsmith* New York: National Bureau of Economic Research. 1954. Pp 120 $1 50

GEORGE GARVY, *Federal Reserve Bank of New York*

THIS study summarizes the substantive finding on one aspect of the savings process. This and other aspects will be considered in full in a forthcoming monograph on the growth of financial intermediaries and their role in the capital markets. The monograph will be part of the project on capital formation and financing sponsored by the Life Insurance Association of America. Several *Occasional Papers* stemming from this same research project (perhaps the most important undertaking of the National Bureau of Economic Research since the war) have been published already.

Goldsmith's study presents new estimates for selected years of aggregate assets held by specific groups of financial intermediaries, the shares of these groups of intermediaries in the main types of assets and liabilities, and the share of financial intermediaries in national wealth and assets One chapter is devoted to each of these topics A discussion of the nature and the limitations of the study and a concise summary of the findings complete the report

Goldsmith's paper is essentially limited to the presentation of new statistical material Some of the new series have been developed in connection with the author's three-volume study of *Savings in the United States,* now in press Since fragmentary basic data preclude the preparation of detailed continuous annual series for the most important financial intermediaries, Goldsmith has chosen to clarify the main movements by providing estimates for eight benchmark years between 1900 and 1949

The essence of the report is in the 27 tables Seven charts depict salient changes over the fifty years. The text in the main describes changes as they unfold between the consecutive years selected, the emphasis is on the what, not on the why. A provocative interpretation of the factual material is, however, contained in the introduction by Simon Kuznets, who was in charge of the over-all project and who endeavors to indicate the place of Goldsmith's findings in the cooperative project of which it forms a part.

The present study provides for the first time comprehensive statistical evaluation of the role of financial intermediaries in external capital formation since the turn of the century. The evaluation of the statistical spade work underlying the various tables presented in this report must await the publication of the full monograph (one can only guess the amount of ingenuity and statistical skill that must have been involved in fitting the pieces to-

gether); so must our curiosity concerning the author's analysis and interpretation of the role of financial intermediaries in the saving process.

In the meantime, two questions may be raised, one about the concept of financial intermediaries and the other about the grouping of the various intermediaries. The author does not discuss why he has made the concept of financial intermediaries so broad as to include social security funds, the Federal Reserve System, and the various government lending agencies, he merely lists in a footnote the institutions included. The justification of the choice made must await publication of the monograph. But the concept used is certainly broader than suggested in Kuznets' introduction, where financial intermediaries are defined as "institutions engaged in investing funds mobilized from a large number of individual and other savers." Views may differ as to whether all institutions holding intangible assets or only those which are repositories of savings should be classified as financial intermediaries And surely, the "mobilization of savings" through social security reserve funds differs in many important respects (including their availability to meet deficits in the private sectors of the economy) from the accumulation of assets in savings institutions.

Fortunately, detail given even in this condensed preliminary report makes it possible for the analyst who prefers a narrower definition or different groupings of financial intermediaries to derive alternative aggregates and percentage distributions by simple arithmetical computations the building blocks are there To facilitate analysis, the basic data are organized on a uniform plan thus, the same twenty categories are included in the stubs of all tables in the chapter on the share of financial intermediaries in various types of assets even though the lines for Federal Reserve Banks, the postal savings system, and government lending institutions show blanks except in the table on holdings of government securities

Treating government institutions (including such government lending agencies as the Reconstruction Finance Corporation and the Home Owners Loan Corporation, which obtained practically all their funds from the Treasury) as just another type of financial intermediary is, perhaps, the most debatable aspect of the study previewed in Goldsmith's report The author by no means ignores the problem, brief reference is made (p. 102) to a different grouping of financial intermediaries into private and public sectors. Nearly the entire growth in the ratio of assets of financial intermediaries to national assets (the most comprehensive among the several ratios used to measure their relative importance) between 1929 and 1949 reflects the growth of public intermediaries The growth of public intermediaries (from 1 to 5 per cent) reflects essentially the monetization of public debt by the Federal Reserve System and the growth of social security (mainly OASI) reserve funds; in the thirties, the substitution of government credit for private credit (by acquisition of mortgages, preferred stock and other assets) by government lending agencies was another important factor.

The present report serves well the double aim pursued by the National Bureau of Economic Research in its *Occasional Papers*. to make available to students as promptly as possible new statistical data and the most important substantive findings by short-cutting the delays involved in the publication of substantial monographs, and to whet the appetite for the main course.

The Volume of Residential Construction, 1889–1950 Technical Paper Number 9, Studies in Capital Formation and Financing *David M. Blank* New York. National Bureau of Economic Research 1954 Pp 99 $1 50

ALBERT H SCHAAF, *University of California (Berkeley)*

THE primary purpose of Mr. Blank's monograph is to describe the methods used in a new estimate of residential construction for the years 1889 to 1920. The estimate was prepared as a part of a larger study on the financing and formation of capital in residential construction which is in turn to be a part of a National Bureau of Economic Research project on long-term capital accumulation and financing in the United States

The new construction series are based on hitherto unused data from a Works Progress Administration project which gathered information on building permits from local records They are linked with Bureau of Labor Statistics figures for 1920–1950 to provide a complete picture from 1889 on. The WPA records provided data for 1920–1929 but, after some consideration, the current BLS estimates for this period were retained

Except for a brief discussion in Section 2, the monograph is devoted to a description of the problems encountered in making the new estimates and the manner in which they were resolved. On the whole, it appears to be a very workmanlike job Of considerable interest is the comparison of the new series with older ones by Chawner, Long, Wickens, and Colean. The monograph provides good grounds for the argument that the new estimates are considerably more reliable They generally show from 10 per cent to 20 per cent fewer private non-farm housekeeping dwelling unit starts than the older series for the years 1900–1920 and about 10 per cent more during the decade of the 1890's

The monograph also presents new estimates of non-housekeeping dwelling unit starts, 1891–1914, and of expenditures for both housekeeping and non-housekeeping units A decline of about 40 per cent in the index (in 1929 prices) of average expenditure per unit between 1890 and 1950 is noted Although the brevity of the treatment may have necessitated the absence of any qualification for this finding, it seems well to point out that conclusions based on such a figure are apt to be highly tenuous. All researchers using index numbers are well aware of the complications involved in comparisons over long periods of time and certainly a dwelling unit built in 1890 was a rather different product from its 1950 counterpart.

Econométrie, Colloques Internationaux du Centre National de la Recherche Scientifique XL, Paris, *Editions du Centre de la Recherche Scientifique*, 1953. Pp v, 332.

R. W. PFOUTS, *University of North Carolina*

THIS star-studded, French presentation offers the reader a stimulating series of essays but confronts the reviewer with an extremely difficult task. The reviewer's problems arise from the diversity of the topics covered by the essays, yet the essays have a central theme, the foundations and applications of the theory of risk in econometrics. This was the theme of a conference held in Paris in 1952, and the essays were the contributions of the participants in the conference.

The conference brought together a most distinguished group The American representatives were (in order of appearance in the volume) Savage, Arrow, Friedman, Samuelson and Marschak. The European participants (arranged in the same manner) were Guilbaud, de Finetti, Allais, Wold, Massé, Morlat, Frisch, Boiteux, Ville, and Van Dantzig Other distinguished participants made brief comments on the major contributions of those listed.

The pervasiveness of risk in economic life is indicated by the fact that risk was considered in connection with such diverse topics as stocks, income distribution, organization, electric costs and tariffs, and credit But in spite of the interest and relevance of these topics, the group devoted more attention to subjective probability and the theory of cardinal utility in the manner of Von Neumann and Morgenstern than to any other single topic.

A large part of the discussion of utility and probabilities is devoted to an examination of the axioms of the Neumann-Morgenstern development of utility. Quite understandably references are made frequently to the work of Marschak, Samuelson, and Savage that clarified and extended the Neumann-Morgenstern postulates that support the thesis of Bernoulli In cases where the attainment of alternative sums of money or alternative bundles of goods is not certain but depends on probabilities, Bernoulli believed that one should behave in such a way as to obtain a maximum expectation of utility from the money or the goods, that one should maximize his "moral expectation."

A good deal of the discussion of utility centers on the assumption that Samuelson has called the axiom of strong independence. This axiom states that if A and B are prospects between which the player is indifferent, C is a third prospect and p is a probability, then $[pA+(1-p)C]$ is indifferent to $[pB+(1-p)C]$ From the standpoint of economic theory this statement appears to lack generality because of the possibility that A and C, say, are connected by a complementarity relation stronger than that connecting B and C. But from the viewpoint of probability theory, it is elementary that the axiom is acceptable because the prospects involved are mutually exclusive and the possible complementarity will never in fact be realized.

Nevertheless the strong independence ~~axiom is~~, at first acquaintance, likely to bring out the schizophrenic in the reader as he views it first from one side

and then from the other before finally separating the theory of utility under conditions of certainty of obtaining the desired goods from the stochastic case in which probabilities of obtaining the goods must be considered. Samuelson shows that strong independence is a necessary assumption for the Bernoullian theory.

The most vocal opponent of the neo-Bernoullians was Allais. He attacked the "American School" from so many different directions that it is almost impossible to give a brief summary of his objections other than to report that his attitude is one of total rejection. But it may be said that Allais evidently felt that the basic error of his opponents lies in their neglect of the dispersion of the psychological values of the various quantities involved.

Allais had at hand a substitute for the Bernoullian theory. He would resort to the Weber-Fechner method of minimum perceptible differences. On this basis he obtains a theoretical preference function that has certain desirable properties. To the reviewer this seemed to involve the danger of an unnecessary retreat toward old fashioned notions of measurable utility.

The reader who gets an unreal feeling from discussions of subjective probability and from reading about a world in which a consumer apparently must weigh a sixty per cent probability of a dozen shirts against a forty per cent probability of a boxer puppy in establishing a preference pattern can draw comfort from the volume under review The authors were well aware of the unreal aspects of the Bernoullian theory, and, as indicated, there is considerable controversy over the validity of the entire approach. There are many refreshing references to "real men" in the volume especially in the discussions of the major papers

Probably the element of risk is less important in the theory of utility than in any of the other topics discussed If this is true, it seems unfortunate that more attention was not given to some of the other economic applications of the theory of risk Nevertheless, some of the other applications are discussed and the discussion of utility goes a long way toward clarifying the Bernoullian theory. Probably one unconsciously discounts the importance of the discussion of utility because the book appeared after the symposium featuring Wold, Shackle, Savage, Manne, Charnes, Samuelson, and Malinvaud that appeared in *Econometrica*.[1] Actually the Paris Conference preceded the appearance of this symposium, and, no doubt, influenced the views expressed in the symposium.

This book probably will not receive as much attention as it deserves from English-speaking economists and statisticians because it is written in French. This gloomy prospect is offset to some extent by the fact that a good part of the material has already been expressed in various articles in English.

A brief but helpful summary of the proceedings written in English by Frisch appears near the end of the volume Readers who try this sampler may find it sufficiently challenging to cause them to refurbish their "reading knowledge" of French and try the essays themselves

[1] *Econometrica*, 20 (1952), 661–79.

PUBLICATIONS RECEIVED

Adult Education Association of the U.S.A. *Financing Adult Education in America's Public Schools and Community Councils, Complete Report.* Washington, D. C: National Commission on Adult Education Finance, 1954. Paper. $0 25

———. *Financing Adult Education in America's Public Schools, Summary Report.* Washington D C: National Commission on Adult Education Finance. 1954. Paper. $0 25

Baker, Helen, and **France, Robert R.** *Centralization and Decentralization in Industrial Relations.* Princeton University: Industrial Relations Section, 1954. $4 00

Fox, L. Prepared on Behalf of the Mathematical Tables Committee of the Royal Society *A Short Table for Bessel Functions of Integer Orders and Large Arguments.* Cambridge: University Press, Published for the Royal Society, 1954 Paper 6 s.

Harris, Seymour. *John Maynard Keynes—Economist and Policy Maker* New York: Charles Scribner's Sons, 1955. $3 00.

Lester, Richard A. *Hiring Practices and Labor Competition.* Princeton University: Industrial Relations Section, 1954 Paper $2 50

National Bureau of Standards. *Table of the Gamma Function for Complex Arguments* Applied Mathematics Series 34. Washington, D C: U S Government Printing Office, 1954 $2 00.

Patterson, Gardner, Gunn, John M., Jr., and **Swerdlove, Dorothy L.,** Assisted by **Mary B. Fernholz** *Survey of United States International Finance, 1953* Princeton: University Press, 1954 Paper $2 75

Phelps, Clyde William. *Financing the Instalment Purchases of the American Family* Baltimore: Commercial Credit Company, 1954 Paper

Pompilj, Guiseppe, and **Napolitani, Diego.** *Piano Degli Esperimenti Ed Elaborazione Probabilistica Dei Risultati Con Particolare Riguardo Alla Sperimentazione in Biologia (Design of Experiments and Probabilistic Elaboration of Results, Especially with Regard to Biological Experimentation)* Supplemento A, La Ricerca Scientifica 1954 Paper Lire 2000

Report of the Mid-Century Conference on Resources for the Future. *The Nation Looks at its Resources* Washington, D C: Resources for the Future, 1954 Paper. $5 00

Rolph, Earl R. *The Theory of Fiscal Economics* Berkeley. University of California Press, 1954 $4 50

Societa Italiana Di Statistica. *Atti Della XI e XII Riunione Scientifica, Roma, Aprile 1951 e Febbraio 1952.* Roma Piazza S. Bernardo

United Nations, Department of Economic Affairs. *Economic Survey of Latin America, 1953.* New York: Columbia University Press, 1954. Paper $2 50

——— *Summary of Recent Economic Developments in Africa, 1952–53.* New York. Columbia University Press. 1954 Paper $0 80

——— *Summary of Recent Economic Developments in the Middle East, 1952–53* New York: Columbia University Press, 1954 Paper $1 25

——— *World Economic Report, 1952–53* New York: Columbia University Press, 1954 Paper $1 75

United Nations, Educational, Social and Cultural Organization. *International Bibliography of Political Science* Paris. Imprimerie Union, 1954 Paper $3 00, 850 fr

——— *The University Teaching of Social Sciences Economics* Amsterdam. Drukkerij Holland N V., 1954. Paper. $2 00, 550 fr

United Nations, Interim Co-ordinating Committee for International Commodity Arrangements. *Review of International Commodity Problems, 1953.* New York: Columbia University Press, 1954 Paper $1 00

AMERICAN STATISTICAL ASSOCIATION

REPORT OF THE BOARD OF DIRECTORS, 1954

The Board is happy to announce that membership rose to a new high in 1954 and that it was a successful year financially The ASA's continued promotion campaigns among other learned societies, nominations from ASA members, and the reduction in dues to residents outside North America have brought in more new members than in any other year since 1947 As a result, membership dues have risen to a new high, and this, combined with increases in other sources of income, has had the effect of producing a surplus greater than that originally budgeted. Thus another accrual to surplus has been achieved Details of membership and finances will be presented in the report of the Secretary-Treasurer.

Journal of ASA and the American Statistician

The *Journal of the American Statistical Association* has been steadily increasing in size for the past five years This is revealed by a comparison between the 1949 volume and the 1954 volume In 1949 the total number of pages of the *Journal* was 590, whereas the total number of pages in the 1954 volume comes to 934, a rise of about 60% The increase in funds for the *Journal*, as voted by the Board, has meant more articles per issue, as well as the publication of abstracts of papers presented at the Annual Meetings The practice of publishing the abstracts is now in its third year (those from the 1954 meeting appear in this issue)

The American Statistician has also increased in the average number of pages per issue in 1954 This has provided more space for information about Association activities, as well as more articles, news of interest to members and other features.

New Publications

The new monograph "Statistical Problems of the Kinsey Report" was completed this year and is now available This important work may be purchased by members at a special price substantially below the price to nonmembers. The monograph is the result of a study made at the invitation of the National Research Council, which is sponsoring the work of Dr Kinsey and his associates A committee composed of William G Cochran, Frederick Mosteller and John W. Tukey studied the statistical methods in Kinsey's first volume, and this monograph is the report of their findings

Also available is the "Proceedings of the Business and Economic Statistics Section," a collection of the papers presented under the Section's sponsorship at the 114th Annual Meeting in Montreal This volume is being sold at a very low price to members It is hoped that the "Proceedings" volume can be published each year but this will depend upon the reception of the first issue.

The newest edition of the Membership Directory was also published in 1954. It contains approximately fourteen per cent more names than the previous edition which appeared in 1951 and has a geographic breakdown and a listing of the members according to the Section of ASA in which they are interested The ASA Constitution, the charters of the sections, and other information about the Association are also printed in the new Directory.

Chapters

During the past year the Board granted charters to two new Chapters of the Association—State College, Pennsylvania and the Statistics Section of the Virginia Academy of Science. These bring the total number of active chapters to 32. Inquiry about forming chapters has come from four more areas and 1955 will undoubtedly see a further increase in the number of chapters.

As a result of the agreement between ASA and the United States Employment Service on handling placement for members, each chapter has appointed a person to act as liaison between the chapter and the local USES office. This person will provide the USES with information and advice on the technical aspects of the various fields in which statisticians work, in order to assist the Employment Service in the placement of statisticians

Sections and Committees

A charter has been granted to the Section on Physical and Engineering Sciences (formerly the Committee on Statistics in the Physical Sciences), which brings to five the number of Sections All five Sections are very active in the formulation and presentation of sessions at the Annual Meetings This provides in the program a wide variety of topics of interest to members

The Committee on Monographs and Occasional Papers was established in 1954 for the purpose of providing a body to implement and review additions to the Association's publishing program. It is expected that this Committee will provide an additional impetus to the ever-widening activities of the Association

The Board in 1954 reviewed thoroughly the functions, scope and composition of the Advisory Committee to the Bureau of the Budget and made certain changes which will enhance the value of the Committee to the Bureau. The number of members on the Committee was increased from six to nine with the specification that a majority of the members must be Presidents, Presidents-elect or Past-Presidents of the Association The remaining members shall be present or past officers of the Association The Board also gave greater freedom to the Committee in its capacity as an advisory body to the Bureau of the Budget and to other governmental departments which may request its assistance The complete report of the Board's review of this Committee appears in the Minutes of the Board of Directors Meeting of May 3, 1954.

Annual Meeting

This year, for the first time in many years, the Annual Meeting of the Association was held at a time other than Christmas week. The meeting held in Montreal in September 1954 was the first that was ever held outside of the United States. The shift from December to September was somewhat of an experiment and was made in response to many requests from members that the Annual Meeting be held at some time other than Christmas week. The attendance at the Montreal meeting was quite good and another September meeting is scheduled for 1956 in Detroit. (The 1955 meeting will be held in December in New York City.) A well-balanced program was presented at the 1954 meeting under the chairmanship of Besse B. Day. A reception given by the City of Montreal for the members was a highlight of the Convention

The Board has decided to poll the members on their preference for the time of the 1957 meeting, which is still open. The questionnaire will be sent early in 1955

and the announcement of the time chosen will be made in an early issue of *The American Statistician*. The questionnaire will present three choices: Spring, Fall or Christmas week.

Future Annual Meetings of the Association are planned as follows:

1955—New York City, December 27–30
1956—Detroit, Mich., September 7–10, joint with the American Sociological Society
1957—Open
1958—Chicago, Ill., December 27–30

Annual Council Meeting

At its meeting in September 1954, the Board voted to have the annual Council Meeting at a separate time from the Annual Meeting of the Association. The Council Meeting will take place in Philadelphia on January 7, 1955. For the first time the Board has authorized the use of proxies, either in writing or via a colleague. One of the reasons for this is to provide Board and Council members from the western part of the country with an opportunity to present their views, even if they are unable to attend in person. The proxy has all privileges except a vote.

In order to provide greater rapport between the National Association and its chapters, as well as providing some measure of recognition for the work of chapter officials, the Board has issued an invitation to chapter presidents to attend the Council Meeting in Philadelphia. This will provide the chapter with an opportunity of bringing to the attention of the Board and Council specific problems. It will also give the chapter presidents an understanding of how the executive body of the Association functions.

Regional Meetings

Three Regional Meetings were held during the past year. In April 1954 the Chicago Chapter, in cooperation with other area chapters, held a Mid-Western Regional Conference which was very successful. A Proceedings volume of the Conference was published and copies may be purchased from the Chicago Chapter.

The New York Chapter sponsored a two-day meeting in connection with the celebration of the 200th Anniversary of Columbia University in May 1954.

The West Coast Chapter organised a Western Regional Meeting in Berkeley, California in December 1954 in conjunction with the American Association for the Advancement of Science and other societies.

Present plans for 1955 call for a conference on Business Statistics to be sponsored by the Business and Economic Statistics Section and the Wharton School of the University of Pennsylvania to be held in the summer. The Section on Physical and Engineering Sciences will hold a conference in May 1955 in conjunction with the centennial celebration of the College of Engineering of New York University.

REPORT OF THE SECRETARY-TREASURER FOR 1954

The 1954 promotion campaign, which was directed primarily to members of the Royal Statistical Society, the International Statistical Institute and the Inter-American Statistical Institute, proved very successful. The number of new mem-

bers from other countries more than offset the small loss in membership income resulting from the reduction of dues to residents outside North America The lowering of dues for persons who have difficulty in obtaining dollar exchange has made it much easier for them to obtain the Association's publications

At the beginning 1954, the membership of ASA was 4,900 The number of new members for 1954 is approximately 700, added to this figure are 36 others who reinstated their membership At the end of 1954, about 480 members were dropped because of resignation, death, or nonpayment of dues. Thus, the net increase in membership for 1954 comes to 250, and the Association begins 1955 with a total of 5,150 members Further promotion campaigns in 1955 are expected to continue the growth of ASA

Subscriptions of libraries, business firms, etc , to the *Journal of the American Statistical Association* have also been increasing. The number for 1954 is 1,437, as compared with 1,356 in 1953 and 1,248 in 1952 An increase in subscriptions to *The American Statistician* has also been noted.

The Secretary's office is happy to announce that almost all copies of the symposium, "Acceptance Sampling," which was published in 1950, have been sold The success of this undertaking, the first monograph put out by the Association in twenty years, has made it possible to underwrite other monographs, such as the "Statistical Problems of the Kinsey Report" and the "Proceedings of the Business and Economic Statistics Section "

The Financial Report, which is attached, shows that the budgeted surplus for 1954 will be exceeded by approximately $1,600 Total income for 1954 was budgeted at $54,775, whereas the final amount for the full year will be about $59,250. This increase is due primarily to income from dues and from the sale of publications, although there were increases in most other items of income, as well. Total 1954 expense was budgeted at $52,465; the actual final figure is expected to approximate $55,260 The cost of publications and the expense of the Annual Meeting account for most of the increase in the total expense, with slight increases in a number of other items of expense

As a result of the 1954 accrual to surplus, the Association begins 1955 with approximately $30,000 in total surplus funds Proposed income for 1955 is budgeted at $65,810, while expense is calculated at $63,270. This leaves about $2,500 for addition to surplus at the end of 1955

The practice of providing monographs and other nonperiodical publications at a reduced price to members has enabled the Association to provide a widespread distribution at the lowest cost and it is hoped that it will be possible to continue this practice Income from the sale of publications to nonmembers at a slightly higher price will provide part of the funds necessary for printing future volumes in the expanding publication program

March 30, 1954

To the Board of Directors of
the American Statistical Association:

I have examined the accompanying financial statements of the American Statistical Association relating to the year ended December 31, 1954 My examination was made in accordance with generally accepted auditing standards and, accordingly, included such tests of the accounting records and other auditing procedures as were considered necessary in the circumstances.

The recorded cash receipts for the year were traced to the deposits shown on the bank statements, and the amounts for dues and subscriptions were tested against the membership and subscription records. The paid checks were inspected and related vouchers tested in support of cash disbursements for the year. The bank balances were reconciled with certificates obtained directly from the depositaries, and the cash on hand was counted and reconciled with the books during the course of the examination I did not check the membership and subscription records in detail or make any independent verification of the inventory of back numbers of *Journals*, the office records of which are based, in part, on data assembled in prior years.

In accordance with a resolution passed by the Board of Directors, the expense incurred in publishing a directory, distribution to the membership beginning in 1954 is to be spread over a three-year period although such costs would appear to be applicable primarily to the year 1954. The accounts for the year ended December 31, 1954, reflect a charge of $1,597.24, representing the allocated portion of the directory expense applicable to that period.

In my opinion, the accompanying statements present fairly the financial position of the American Statistical Association on December 31, 1954, and the results of its operations for the year then ended, in accordance with generally accepted accounting principles, except as mentioned in the previous paragraph, applied on a basis consistent with that of the preceding year.

JAMES G JESTER

AMERICAN STATISTICAL ASSOCIATION
BALANCE SHEET

Assets

	December 31, 1954	*December 31, 1953*
Cash in banks and on hand (see note)	$59,741 20	$52,431 05
Accounts receivable	894.71	2,783.14
Investment in United States Savings Bonds, Series G, due 1962, at cost	3,100 00	3,100 00
Inventory of old *Journals*, at approximate cost	2,360 99	2,137 69
Inventory of Monograph on Acceptance Sampling, at cost	45 49	129 24
Inventory of Emblems, at cost	361 50	415 50
Inventory of Monograph on Kinsey Report, at cost	4,482.00	
Furniture and fixtures, at cost less accumulated depreciation	1,908 99	2,088 78
Deferred Charges:		
Deferred Membership Directory expense	3,604 48	
Other	1,136 75	945 55
Total Assets	$77,636 11	$64,030.95

Liabilities and Net Worth

Accounts payable	$17,488.68	$10,430.60
Deferred income (collections applicable to subsequent years):		
Dues	$18,987.00	$16,827.00
Subscriptions	6,535 80	5,817.77
Other	466.84	466.84
Total deferred income	$25,989.64	$23,111 61
Net worth:		
Life Membership reserve	$ 3,796.98	$ 3,579.92
Surplus, per statement	30,360 81	26,908.82
Total net worth	$34,157 79	$30,488.74
Total Liabilities and Net Worth	$77,636.11	$64,030.95

Note: The amount listed as cash in the banks and on hand for 1954 breaks down as follows:

Checking Account (American Security & Trust Co.)	$17,865.34
Hyattsville Building Assn.	1,013 33
Jefferson Federal Savings & Loan Assn	10,561 33
National Permanent Building Assn	5,073 00
Liberty Building Assn.	10,195 26
American Building Assn.	5,117.16
Interstate Building Assn	9,915 78
	$59,741.20

AMERICAN STATISTICAL ASSOCIATION
STATEMENT OF INCOME AND SURPLUS ACCOUNTS

	Year ended December 31, 1954	1953
Income:		
Dues—Current year	$40,232 75	$38,607.00
—Prior year	155.00	852.00
Subscriptions—*Journal*	10,869 22	10,134.80
—*American Statistician*	462.75	443 68
Advertising—*Journal*	1,319.82	1,415.99
—*American Statistician*	125.00	263.97
Sales—*Journal*	2,870 95	1,937.13
—*American Statistician*	146.95	148.02
—Acceptance Sampling Monograph	175.00	302.05
—Emblems, less cost of sales	26 75	83.28
—Biometrics		568 37
—Other	7.41	45 00
Mailing list income	518 92	911.11
Interest income	1,230.70	911 84
Annual meeting	1,047 84	1,559 52
Reimbursement of overhead expenses Bureau of Mines Project		2,207 06
Miscellaneous income	20.43	54.12

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

The Editors welcome the submission of manuscripts for possible publication. They should be typewritten entirely double-spaced, including footnotes, and two copies should be sent to the Editor, W Allen Wallis, 207 Haskell Hall, University of Chicago, Chicago 37. Books for review should be sent to the same address. Unsolicited book reviews are not accepted, but suggestions of titles for review are welcome.

Errata: Readers and authors are urged to submit to the Editor notices of errors found in this or any previous volume. These will be published once a year, in the December issue.

A cumulative Index to Volumes 1–34, 1888–1939, may be obtained from the Office of the Secretary of the American Statistical Association
Reprints of all articles published since 1945 may be purchased from the office of the secretary. Articles which have appeared during the past twelve months are usually available in quantity.

EDITORIAL COLLABORATORS

MEYER DWASS
Northwestern University
ROBERT EISNER
Northwestern University
GEORGE L. EDGETT
Queens University
HOPE T. ELDRIDGE
United Nations
BENJAMIN EPSTEIN
Wayne University
MORDECAI EZEKIEL
United Nations
HELEN C. FARNSWORTH
Stanford University
LEON FESTINGER
University of Minnesota
RICHARD J. FOOTE
Bureau of Agricultural Economics
LESTER R. FRANKEL
Alfred Politz Research, Inc
MILTON FRIEDMAN
University of Chicago
IRWIN FRIEND
University of Pennsylvania
WALTER GALENSON
Harvard University
ERWIN A GAUMNITZ
University of Wisconsin
DAVID V. GLASS
University of London
LEO A GOODMAN
University of Chicago
EUGENE GRANT
Stanford University
FRANK E GRUBBS
Aberdeen Proving Grounds
P M. GRUNDY
Rothamsted Experiment Station
MARGARET GURNEY
Bureau of the Census
MARGARET J HAGOOD
Bureau of Agricultural Economics
MARGARET HAMILL
Ordnance Corps
MORRIS H. HANSEN
Bureau of the Census
ARNOLD HARBERGER
University of Chicago
HARRY H HARMAN
The RAND Corporation
H. O. HARTLEY
Iowa State College
MILLARD W HASTAY
National Bureau of Economic Research
WILLIS L. HASTY
Camp Detrick, Maryland
J. L HODGES, JR.
University of California (Berkeley)
WASSILY HOEFFDING
University of North Carolina
ROBERT HOOKE
Princeton University
DANIEL G HORVITZ
University of Pittsburgh
EARL E HOUSEMAN
Bureau of Agricultural Economics
HARRY M HUGHES
University of California (Berkeley)
RHEEM F. JARRETT
University of California (Berkeley)
GEORGE JASZI
Department of Commerce
EMIL H JEBE
Iowa State College
R J. JESSEN
Iowa State College
OSCAR KEMPTHORNE
Iowa State College
DAVID KENDALL
University of Oxford
MAURICE G KENDALL
University of London
FRANK L KIDNER
University of California (Berkeley)
JACK KIEFER
Cornell University
EDGAR P KING
Eli Lilly and Company
DUDLEY KIRK
The Population Council
CLYDE V KISER
Milbank Memorial Fund
LESLIE KISH
University of Michigan
LAWRENCE R KLEIN
University of Oxford
LLOYD A. KNOWLER
University of Iowa
ALBERT R KOCH
Federal Reserve System
H S KONIJN
University of California (Berkeley)
CARL KOSSACK
Purdue University
WILLIAM H KRUSKAL
University of Chicago
ERNEST KURNOW
New York University
STANLEY LEBERGOTT
Bureau of the Budget
IVAN M. LEE
University of California (Berkeley)
J. M LETICHE
University of California (Berkeley)
JOSEPH LEV
New York State Department of Civil Service
GERALD J LIEBERMAN
Stanford University
JULIUS LIEBLEIN
National Bureau of Standards

Abraham M. Lilienfeld
Roswell Park Memorial Institute
E. F. Lindquist
University of Iowa
Richard Link
Sandia Corporation
Sebastian Littauer
Columbia University
Clarence E Long
Johns Hopkins University
E. Lord
British Cotton Industry Research Board
Eugene Lukacs
National Bureau of Standards
Sherman J Maisel
University of California (Berkeley)
John Mandel
National Bureau of Standards
Eli S. Marks
National Opinion Research Center
Frank J Massey, Jr
University of Oregon
W. Parker Mauldin
Bureau of the Census
Quinn McNemar
Stanford University
Perry Meyers
Perry Meyers, Inc.
Don I. Mittelman
Diamond Ordnance Fuze Laboratory
Geoffrey H Moore
National Bureau of Economic Research
P G Moore
University of London
P. A. P Moran
Australian National University
J. E Morton
National Science Foundation
Jack Moshman
Bell Telephone Laboratory
Frederick Mosteller
Harvard University
K. R. Nair
Forest Research Institute
John Neter
Syracuse University
Gottfried Noether
Boston University
Edward B. Olds
Social Planning Council of St Louis
Edwin G Olds
Carnegie Institute of Technology
Gladys Palmer
Wharton School of Finance and Commerce
Walter Palmer
Young and Rubicam, Ltd.
Edward Paulson
Queens College (N.Y.)
Eugene W. Pike
Raytheon Manufacturing Company
Irvin Pollin
Ordnance Corps
W. H Poole
Young and Rubicam, Ltd.
John Pratt
University of Chicago
Joseph Putter
University of California (Berkeley)
Howard Raiffa
Columbia University
Stanley Reiter
Purdue University
Paul R Rider
Wright Patterson Air Force Base
John Riordan
Bell Telephone Laboratories
Murray Rosenblatt
University of Chicago
Gideon Rosenbluth
Queen's University (Canada)
Alan Ross
Iowa State College
H L Royden
Stanford University
Richard Ruggles
Yale University
Leonard J Savage
University of Chicago
Henry Scheffé
University of California (Berkeley)
Harry Schwartz
The New York Times
Jacob S Siegel
Bureau of the Census
Herbert A Simon
Carnegie Institute of Technology
Rosedith Sitgreaves
Columbia University
Milton Sobel
Bell Telephone Laboratories
Robert M Solow
Massachusetts Institute of Technology
Paul N. Somerville
Virginia Polytechnic Institute
R. Clay Sprowls
University of California (Los Angeles)
William A Spurr
Stanford University
Joseph Steinberg
Bureau of the Census
J. S Stock
Alfred Politz Research, Inc.
J R N Stone
University of Cambridge
Alan Stuart
London School of Economics
G. H Symonds
Esso Standard Oil Company

'ICTOR VON SZELISKI
U. S. Economics Corporation
CONRAD TAEUBER
Bureau of the Census
D. TEICHROEW
National Cash Register Company
BENJAMIN J TEPPING
Bureau of the Census
DONOVAN J THOMPSON
University of Pittsburgh
DONALD R TRUAX
Stanford University
JOHN W. TUKEY
Princeton University
DAVID VALINSKY
City College of New York
UPERT VANCE
University of North Carolina

RUTLEDGE VINING
University of Virginia
HOWARD E WAHLERT
New York University
DAVID L WALLACE
University of Chicago
JOHN E WALSH
Lockheed Aircraft Corporation
LIONEL WEISS
University of Virginia
B L WELCH
University of Leeds
P. K WHELPTON
Miami University
HUGH H WOLFENDEN
Minneapolis, Minnesota
F. YATES
Rothamsted Experiment Station

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 271 SEPTEMBER 1955 *Volume 50*

STATISTICS AND OBJECTIVE ECONOMICS*

Morris A Copeland
Cornell University

One could argue that the history of economic thought has long been shaped—at least in part—by statistics. In the nineteenth century value theory was strongly emphasized, the theory of production rather slighted, at that time price statistics were far more plentiful than production statistics Again laissez-faire was bolstered by a price, wage, and interest theory that largely ignored the government as a customer, as an employer, and as a borrower, at that time the government budget was small in relation to national income Further, international economics emerged as a special field, international trade statistics and foreign exchange rates loomed large among our earlier economic time series.

This is by no means all there is to the earlier influences of statistics on economics. But the major statistical impact has come since World War I, and it is that impact we are concerned with here

Let me begin by considering the nature of the stimulus that has been applied to economics Really there are two stimuli On the one hand, there is statistics as a scientific method of making observations and drawing inferences, particularly a method of inferring the characteristics of a population from the characteristics of a sample which is in some sense random Modern statistical method is a widely applicable technique of scientific investigation that embraces the planning of the pattern of observation as well as the mathematical logic of inductive inference.

On the other hand, we have statistics in the sense of the facts of observation themselves, or rather, since we are here concerned with economic statistics, the facts which emerge from what is called statis-

---

* An earlier version of this paper was presented on May 24, 1954 at a symposium sponsored by the American Statistical Association, New York Area Chapter, in connection with the Bicentennial Anniversary of Columbia University.

tical collection and compilation. Statistical collection and compilation necessarily goes farther than mere observation; it involves logical inference and trained judgment. Moreover, this kind of fact finding is commonly beyond the capacity of any one individual. It takes an organization to find the population of the United States, and more than one organization to find the gross national product. Yet such findings of quantitative facts are for the economist the nearest available analogues to the measurements of physical science. Possibly they are much closer analogues than has sometimes been supposed, despite their inferential nature. At all events economic statistics are the empirical measurements the economist has to work with.

Since World War I there has been a very great increase in the stock of economic measurements economists have at their disposal. There has also been a very great development in statistical method. But the impact of statistics on economics has come almost entirely from the measurement side; the greatly improved mathematical-statistical method has made but little impression on the recent course of economic thought.

The reason why statistical measurements have exerted a substantial influence is not far to seek Economists had long aspired to make their subject a genuine science For a time they had looked to Newtonian mechanics as an example of what a science ought to be. The aim was a mechanics of the market place, but the result turned out to be more like Euclid than Newton Presently there were some that argued that geometry, rather than mechanics, was the proper model for economic science But there were others who regarded the so-called laws of economics as too static, too deductive, and too remote from the real world, and who yearned to make economics an objective empirical study of our actual economy.

Wesley Mitchell was a great leader in this latter group. In addressing the American Economic Association as its President, he outlined what he thought quantitative analysis, aided by "the increase of statistical data, the improvement of statistical technique, and the endowment of social research", might do to—and for—economics He predicted that "the men now entering upon careers of research may go far toward establishing economics as a quantitative science". And he anticipated that moving in this direction would require "a recasting of the old problems into new forms amenable to statistical attack". Further, contrasting the Newtonian or mechanical and the statistical conceptions of nature, he said the latter "may be expected to make more radical changes in economics than it makes in physical theory". He com-

mented too on the prospect for the traditional type of economics which he characterized as "deductive" and as involving "excursions into the subjective". He thought it "unlikely that the quantitative workers will retain a keen interest in imaginary individuals coming to imaginary markets with ready-made scales of bid and offer prices Their theories will probably be theories about the relationships among the variables which measure objective processes. There is little likelihood that the old explanations will be refuted . . much likelihood that they will . . . drop out of sight in the work of the quantitative analyst".[1] Mitchell's address was delivered in 1924

Before World War I there had not been very much in the way of statistical measurements on either side of the Atlantic to implement the kind of objective quantitative inquiry Mitchell expected to see develop And in many respects American economists had been less well off than their European colleagues To be sure our Statistical Abstract was a substantial octavo volume, and it was literally then, as it is today, only an abstract But the quantity of statistical measurements is not a matter of weight and tale alone. They must be à propos.

Economic statistics, like other measurements, are by nature comparative A single measurement can be quite useless, usually you need other measurements with which to compare it. In the case of economic statistics there is one type of comparison that has long been recognized as particularly important—comparisons in time. Economic time series, such as the series on our wheat crop each year, are a particularly useful class of economic statistics And within this class there is one subclase that possesses far greater utility than the others—economic tims series that are on a current quarterly or more frequent basis This kind of measurement is essential for analyzing the current business situation Before World War I we had only a thin, scattered assortment of such series, e g , imports and exports, foreign exchange rates, stock, bond, and wholesale commodity prices, bank clearings in leading centers, freight-ton-miles, business failures, pig-iron production[2] There were no comprehensive current figures on employment and unemployment, on inventories, on construction, on retail trade or retail prices, on bank credit. The idea of an index of physical production was widely deemed not feasible. Monthly measures of total personal income would have been thought an idle dream. The gross national product was only a theoretical concept mentioned by Adam Smith and then largely

[1] *American Economic Review*, March 1925, 1–12. There are various other quotations from this address below.

[2] See *Historical Statistics of the United States, 1789–1945*, Washington. Bureau of the Census, 1949, Appendix I.

forgotten It had not yet been given even an annual or occasional statistical significance.

This listing of gaps in the current quarterly and monthly time series available before World War I makes it clear that our area of ignorance was much larger than our area of knowledge In terms of annual, decennial, and other noncurrent compilations we were considerably better off. Such compilations covered agriculture, mining, manufacturing, railroads, banks, government, and several smaller industry sectors. In the decade 1904–13 about three fifths of our national income originated in these sectors Moreover, there were clues to what was going on in the others, notably the clues in the decennial census of occupations But it should be added that the statistical record of our pre-World-War-I economy is substantially more complete today than it was in 1918. A good deal of what we now know has been pieced together in the meantime Piecing together is an essential part of the work of statistical fact finding, and a part that has come to be performed far more adequately in the last thirty odd years

This building up of the retrospect has meant a significant addition to the stock of statistical measurements at our disposal. But most economists would rate as far more significant what has been accomplished by way of building up the current picture With regard to the magnitude of this accomplishment it can be said that we now have what we lacked before, an effective working stock of current time series, a far better one than that possessed by any other country in the free world There are still gaps to be filled, and there are shaky figures that need to be made firm But it is a sufficiently well rounded and reliable stock to make effective empirical economic analysis possible.

The measurements of physical science are mostly *ad hoc* measurements When an investigator needs a measurement he finds a way to make it Scientific curiosity is in general the motive that provides such measurements Scientific curiosity has played a part, too, in making our stock of economic measurements what it is today. But its role in this connection has necessarily been a modest one. I have noted that the retrospect has been improved by a process of piecing data together. This piecing-together process is responsible, also, for a vast improvement in our current statistical picture. Without it that picture would still be a very spotty one. And scientific curiosity has been a main motive in the development of the piecing-together techniques However, without the basic current and less frequent periodic reporting services that have come into being since World War I, there would not be much that could be pieced together. And scientific curiosity has hardly been

the main promoter of these services. Probably the most that can be said is that it has abetted curiosity-with-an-axe-to-grind. But even curiosity-with-an-axe-to-grind has played a modest role. It is true that *ad hoc* collection services provide a substantial part of our basic data. The various censuses, the monthly payroll and employment reports to the Bureau of Labor Statistics, and current trade association reports are cases in point. But there is a much larger body of basic data collected or recorded for some administrative purpose. For example, there are tax returns and the accounting records of businesses and governments. In the vast assortment of basic data we rely on today the statistics that are by-products of administrative record keeping and administrative reports bulk far larger than do those that result from *ad hoc* collections

We will not attempt to explain how our stock of by-product measurements has come to be what it is today Such an inquiry, however intriguing, would be a digression But the comments just made on the role of scientific curiosity in adding to our stock of measurements have been offered because they help to answer a directly pertinent question that has surely occurred to you If the economic measurements accumulated since World War I have exerted a substantial influence on the course of economic thought, because there was a group of economists who yearned to make their subject an empirical science, how comes it that the greatly improved mathematical-statistical method, which now offers to all fields of science a plan of inductive inquiry, has made but little impression?

Let us consider this question under two heads, one relating to the problems of the accuracy of economic measurements, the other, to the problems of constructing economic models

If we think of statistics as a genus of quantitative facts, it would seem that economic statistics should be regarded as a species of that genus. Now the proponents of modern statistical method regard the members of the genus as numerical characterizations of what they call a population or universe, or of a sample of a population or universe Presumably the figure \$365 billion for our gross national product in 1953 is a numerical characterization of a particular universe, the 1953 national product, and since to some extent sample data were used in arriving at this figure, it is natural for the proponents of statistical method to ask, "Why not use the theory of sampling to provide a measure of its accuracy?"

Certainly it is not the practice to say that GNP was \$365 billion $\pm x\%$. In fact national income figures were often published in this

form thirty-odd years ago, but the practice has gone out. Most economic statistics are now published without any attempt to specify the error quantitatively.[3] There are exceptions to this rule, of course. But the rule is against such a specification

The ground for this rule is surely not that the degree of accuracy attained by economic statistics is so high—quite the contrary. One reason for the rule is that the errors modern mathematical statistical method has enabled us to measure are not the only errors to which economic statistics are subject. Often they are relatively unimportant. For one thing there are sampling errors that have a time dimension that is not yet covered by probability theory—I mean errors which arise because a sample-universe relationship derived from study of a benchmark year is applied to sample data for subsequent periods, and in these subsequent periods the sample may have gotten out of line with the universe to an extent not reliably measured by any available statistic. There are conceptual errors too—thus the articulation of the parts of a total like GNP may not be quite correctly designed The articulation of the parts may be imperfect for this reason: It may be imperfect also because the basic data do not fully conform to standard specifications, e.g, the basic data used in putting together a consolidated balance sheet for our banking system may not all refer to exactly the same date; or they may not all define foreign banks in the same way. Further, there is always the possibility of sheer mechanical errors And of course there are incomplete and doubtful data that must be used in some of the steps taken in arriving at a comprehensive total like GNP. Again, the data available are continually changing—probably no two successive annual estimates for a total like GNP are exactly comparable This is not a complete catalogue of sources of error, but it is perhaps enough to indicate why attaching a single percentage error measure derived from a probability calculation to many current economic statistics would only be misleading Instead, in the case of time series, the prevailing practice is to mark some figures *p* and others *r*.

But, as it applies to time series, there is another reason for the rule against showing the percentage error. The main interest here centers not on one absolute error but on various relative errors. What counts is not so much the absolute level of a comprehensive total like disposable personal income but its year-to-year and quarter-to-quarter movements and its relations to other totals, e g , personal consumption ex-

---

[3] Thus none of the current monthly figures regularly appearing in the *Survey of Current Business* is accompanied by such a specification, although many of them are based on sample data.

penditure and personal saving. Attaching a separate percentage error figure to each quarterly estimate of disposable income would not be a satisfactory way to indicate the accuracy of either quarter-to-quarter or year-to-year movements in this series. Nor would separate percentage error figures attached to both disposable income and consumption estimates be too helpful in appraising the accuracy of a measure of the relation between them.

These comments indicate why the scientific error measuring technique that has been elaborated by modern mathematical statistical method has not made much impression in the field of economic statistics They also have an affirmative implication that deserves attention here. Every basic compilation of statistical data and every set of pieced-together economic measurements ought to be accompanied by an adequate descriptive statement. Such a statement is the best available substitute for a quantitative appraisal of errors; but it should serve a still more important purpose too It is needed to tell the user exactly what the figures mean.

This affirmative implication has two edges—it asserts an obligation on the producers of statistics, and a corresponding obligation on those who use them in economic analysis.

The producer has an obligation to provide an adequate description of his product. In the case of those basic data that are by-products of administrative records or reports this obligation may rest on the processor, if the primary producer has not met it properly. And with respect to the processing or piecing-together, the obligation should read like this: Specify what you have done so fully that you could expect others to repeat the process and come out with substantially the same findings. If economists want to aim at something like the kind of objectivity in their measurements that attaches to the measurements of physical and biological science, this is surely the sort of standard they should set for themselves. But it suggests a material qualification on what has been said above about our progress in developing economic measurements. We now have a lot of measurements but with regard to many of them, if we are candid, we must admit that we do not know enough about them to reproduce them. Undoubtedly there has been substantial progress in statistical specification statements, but to conform to the objectivity canon proposed there is still a long way to go And there are pressures that work against going that way very rapidly The preparation of a statistical specification statement is a tedious job that cannot well be delegated to a clerk. Time and money can be saved by slighting it and there are both time and money pres-

sures. Only a fraction of the statistical public will object if the specification is delayed a few years, or if it is quite brief and vague. Very few will object if they cannot repeat the process and confirm the results. However, those few do something to bring about the correction of errors in economic time series And they exert some pressure—perhaps disproportionate to their numbers—toward better specifications

The other edge to the implication from my comments on the nature of the errors in economic measurements relates to the users Those who use such measurements in economic analysis have an obligation to understand the meaning of the figures they use, i e , to know what they are doing No doubt this statement will strike many non-economists as an unnecessary amplification of the obvious But, unfortunately, a kind of division of labor has grown up under which some people make it their business to know in detail how concepts like GNP, personal income, and personal saving are defined, and to play a part in the piecing-together process of constructing economic measurements, others prefer to concentrate on exploring the relationships among these economic variables With such a division of labor, one can contemplate the possibility that Mr X may find a new and better formula for predicting what is currently called "personal saving" from "disposable income" and yet not realize that what he has learned to predict is really not personal saving but a mixture of household saving, institutional saving and additions to noncorporate business surplus, a mixture that does not necessarily reflect all of household saving It can be argued that such an unhappy possibility is a necessary cost of progress in economic inquiry; both the constructing of statistical measurements and economic model building can be very time-consuming occupations But the validity of this argument is open to question Economic model building is still a kind of job that can be done by people who are not full time specialists I think those who concentrate on analyzing relationships among variables whose meanings they have not really stopped to investigate are unnecessarily impatient to get results They may get results by impatience, but for the longer pull the obligation to know about the figures they are using remains.

The other part of the question posed above as to why the recent developments in mathematical statistical method have exerted but little influence on the course of economic thought relates to model-building. Present-day statistical method offers a kind of guidebook to model-builders Why hasn't this guidebook done more to shape the course of economic model-building? One reason is quite simple The guidebook proposes a plan for taking observations. It tells how to sample a uni-

verse so as to get a reliable model for it. Economic measurements are the quantitative observations the economist wants to get, and he particularly wants time series What he wants is not a sample of a time series, but the whole series for a period of years, if he can manage to get it Often he would like it for a longer period than circumstances permit. In taking time-series mesurements he takes all he can get. The guidebook is no help It is true there are areas of economic investigation where *ad hoc* observations can advantageously be made and a plan of observation is needed It is true, too, that there are other fields of inquiry in which the possibilities of *ad hoc* observation are more narrowly confined than in economics, e g , history and paleontology None the less, the part of economics that is not helped by a mathematical statistical plan for taking observations is a substantial part of the whole.

But there is a second and more fundamental reason for the somewhat limited usefulness of the new mathematical-statistical guidebook in economic model building. This reason involves both the nature of the guidebook and the nature of economics. The guidebook aims at general applicability to all fields of inquiry and general applicability implies that the problems of the various fields—say physics, biology, and economics—conform closely to a single pattern We contend that they do not, and that the guidebook, since it has been designed mainly for the physical and biological sciences, is not particularly well designed to serve the purposes of the economist. The economist's inquiry problems differ from those of the physicist and biologist because of two special characteristics of many of the time series with which he has to work These two characteristics apply peculiarly to social and economic statistics. They are of special concern to the economist, because economic time series constitute the vast bulk of social and economic time series. It will be convenient to refer to series in which the two special characteristics are prominent as one-culture time series

Per capita disposable income in the United States at 1947 prices, 1929–53, may be taken as an illustration of a one-culture time series. Let us contrast this type of series with such a series as total annual precipitation at Chicago, 1929–53. The first special characteristic that distinguishes the former is that the geographical specification is more than a mere geographical specification, it is a culture specification as well. And culture specification is important, because cultural differences can pose major obstacles to making interspatial comparisons. Thus while there are doubtless serious difficulties in saying what per capita income in the United States in 1929 would be equal to $1500

in 1950, such a comparison seems quite safe and simple when we face the problem of trying to say how many dollars of per capita income in the United States in 1950 would be equal to a per capita income of £200 in the United Kingdom in that year. But cultural differences do not hamper the making of interspatial comparisons of meteorological or other physical measurements Because the geographical specification is partly a culture specification in the case of many economic measurements, and because one-culture intertemporal comparisons of such measurements are often so much easier and safer than intercountry comparisons, one-culture time series are today the outstandingly important category of economic measurements. Hence, too, economic model builders—to the extent that they have operated empirically—have inevitably devoted most of their attention to time-series or period analysis models.

The second special characteristic of a one-culture time series is that its meaning may be gradually changing as the culture, of which it is an aspect, evolves So, too, if we think of a set of one-culture time series as a set of variables, the relations among these variables may be gradually changing. Because of this characteristic the economic model-builder faces a theoretical dilemma Theoretically he can attempt to develop a period analysis model that will describe the course of cultural evolution Alternatively he can assume that the evolutionary process is sufficiently gradual so that he can get useful results if he ignores it. The first alternative offers no prospect of early success; it is strictly a theoretical alternative in the present state of our understanding.

The second and only practical alternative today is rather awkward. The economic model builder necessarily works with time series covering a finite period and fits his model to the observations for that period. He would like to assume that these observations constitute a representative sample of a longer period, and to use his model to draw inferences about that longer period. But because he is following the second alternative he cannot get out of the past a random sample of a period that includes the future He cannot use the methodological guidebook of mathematical statistics to appraise the validity of such inferences Instead he adopts a rule of thumb to determine the confidence with which a stable relationship among variables during the period for which observations are available can be used to draw inferences about a longer period The rule is that confidence diminishes as the length of the extrapolation period increases

In view of the gradual-evolution characteristic of one-culture time series the rules of mathematical statistical inference can not be applied

to extrapolations. But the economic model-builder can still use various statistical techniques much as if they did. He can and does. In particular he uses mathematical best-fit techniques to determine his parameters and on occasion he postulates probability distributions for his error terms and makes his parameters functions of time. On the other hand there are many rather mundane techniques employed by statistically minded economists to the development of which mathematical-statistical theory has made negligible contributions. Some of these are used in the piecing together process of statistical fact-finding and in designing and operating index numbers (e g, splicing); others in analyzing time series variations (e.g, calendar adjustments). Such mundane techniques may well have done rather more for scientific method in economics than the mathematical-statistical guidebook.

The restrictions on the applicability of mathematical statistical method we have been considering relate to time-series analyses. With respect to cross-section analyses I shall stop only to say that the usefulness of mathematical statistical method varies with the problem. There are problems for which it seems ideally designed. There are problems for which its applicability is even more seriously restricted than for those connected with time series And there are many problems in between.

We have attributed the very limited impact of recent improvements in mathematical-statistical method on the course of economic thought to the fact that that method is not well adapted to the problems of inquiry in the field of economics. The direct impact would, in any case, be limited to those problems which lend themselves to statistical treatment. Even within this area it is further restricted because there is a group of economists engaged mainly in statistical-economic research, who feel modern mathematical-statistical techniques to be malapropos and make a point of avoiding their use. Still, so long as there is a substantial group who do use them, one might well look for a significant impact.

There is, indeed, at least an incipient change in economic thought that can properly be attributed primarily to statistical theory In the older—still widely held—deductive, subjective conception of an economic model the endogenous variables were all uniquely determined. With the error terms that characterize the statistically fitted equations of empirical model analyses some measure of indeterminacy is beginning to replace this older absolute determinism.

So much for the slight impact of statistical method on the course of economic thought. What has been the nature and extent of the impact

of statistical measurements? Have they given economics the sort of objectivity Mitchell had in mind? In considering this question I must in fairness note that I omitted some of the qualifications he attached to his forecast. One of them suggests a kind of obduracy in the deductive, subjective approach. He expected "quantitative analysis (to) produce radical changes in economic theory" but he warned us that it "*does not promise a speedy ending of the types of economic theory to* which we are accustomed". We have had radical changes; the "excursions into the subjective" continue, although the rationalizations it is their purpose to provide have no proper place in the theoretical framework of any true behavioral science[4]

Following Keynes the problems of economic theory have been recast The field is now divided into two main parts, macroeconomics and microeconomics In macroeconomics, quantitative analysis has made great progress Quantitative workers now have at their disposal an impressive array of basic economic concepts which have something like the objectivity that attaches to the concepts of physics, in that they are defined operationally, i e, in terms of the way they are measured. Among these are production (i e, deflated GNP), wealth, personal income, personal consumption expenditure, gross private domestic capital formation, money (i e, the total currency and deposit liabilities of the banking sector to other sectors), bank credit, the wholesale price level, unemployment, the labor force These operationally defined macroeconomic concepts have given economists a set of objective variables whose behavior they can investigate empirically.

Quantitative workers conceive the task of macroeconomic theory to be to develop equations which relate these economic variables, to use these equations as well as a vast amount of detailed facts, to analyze the past behavior of our actual economy as it is reflected in changes in the aggregative variables, and, on the basis of such analyses to form opinions about the future

Consider one of these equations, the consumption function This equation is not as yet a very perfect instrument; while a number of improvements have been made in it in the past decade, much work still remains to be done. In contrasting the mechanical and the statistical conceptions of nature, Mitchell emphasized "imperfect approximation" as characterizing the latter The consumption function is unmistakably an imperfect approximation But quantitative analysts have found it very useful nonetheless. It is what is called a behavioristic equation, i e., it is a statistical best-fit to the available measurements Historically

[4] It would of course be scientifically appropriate to investigate, on the basis of an operational definition of rational behavior, how much of human behavior is rational

it is interesting that when Keynes proposed the consumption function he indulged in subjective analysis, it is interesting, too, that this kind of rationalization of consumer behavior has dropped "out of sight in the work of the quantitative analysts" on improving the consumption function and that these quantitative analysts have introduced lags and stickiness factors commonly slighted by those who stress subjective rationalizations. But these quantitative analysts do not have the field to themselves. And their neglect of the subjective has not gone without protest.

Many a macro-economic theorist who has been accustomed to defining his basic concepts in terms of individual choices and expectations, has continued to prefer such subjective definitions to objective, operational ones Quite possibly he feels the latter to be lacking in logical precision, certainly they smack of "imperfect approximation". And the shift from subjective to operational definitions would require a fundamental change in habits of thought that is not easily made except when one is young

Those who prefer subjective definitions in general prefer a corresponding type of economic analysis Thus there is today a considerable group of economists who see no merit in a purely behavioristic equation like the consumption function They argue that there are "in fact two concepts of propensity to consume". One they characterize as "formal", "aggregate", and "*ex post*" [5] This is the one we have been considering Though they call it *ex post* it has been used in making projections into the future It might, therefore, be better to call it objective The other concept is not amenable to statistical investigation The economists who insist on it distinguish it as "psychological", "individual", and "*ex ante*". They tend to deprecate the objective, statistically determined propensity as "the tautological concept of the marginal propensity to consume",[6] and to say that a theory of business cycles that depends on it reduces to "a definitional proposition of no significance" [7] There is also a slightly milder condemnation of this kind of quantitative cyclical analysis which asserts that it tends "to submerge the process of economic change" in static or "instantaneous pictures".[8]

These reactions are of course reactions to Keynesian model analysis.

[5] Haberler, Gotfried, "Mr Keynes' theory of the 'multiplier'", reprinted in *Readings in Business Cycle Theory*, Philadelphia Blakiston, 1944, 193–202

[6] Machlup, Fritz, "Period analysis and the multiplier theory", reprinted in *Readings in Business Cycle Theory*, Philadelphia Blakiston, 1944, 203–18

[7] Fellner, William, "Employment Theory and Business Cycles", Section 6, in *A Survey of Contemporary Economics*, Philadelphia Blakiston, 1948, 53–5

[8] Williams, John H, "An appraisal of Keynesian economics", *American Economic Review*, Supplement, May, 1948, 288–9

Those who have labored to construct and to improve a statistical model along Keynesian lines are quantitative analysts. And the criticisms of this type of analysis just cited are in effect criticisms of any purely behavioristic analysis, Keynesian or otherwise. Those who bring the charges of tautology and static pictures appear to hold that a theory of the cycle must be something more than an hypothesis that approximately describes or fits the observed facts and that can be used in making projections; that it must indeed explain or rationalize what happens in terms of hypothetical individual expectations and preferences. On this ground Mitchell's own hypothesis for the cycle was long ago crititicized as not a theory. Of course this was mere name-calling; not an intellectual criticism. But the recent critics of behavioristic quantitative economic analysis would have been better advised to say "not a theory" than to say "tautology" or "static". They are vulnerable on the truism count themselves Besides, the charges of tautology and statics cannot be sustained against a Keynesian statistical model. Whatever else may be said about this type of model it is not a definitional proposition devoid of empirical significance and not a static picture. No tautology could possibly lead logically to a wrong factual conclusion as did the model which produced the VJ Day forecast; nor could a static model It takes a period analysis model to produce a prediction.

Perhaps these comments on behavioristic quantitative analysis and its critics suggest that economic theorists can be classified as either behaviorists or subjectivists; they cannot There are not a great many quantitiative analysts today who are—as Mitchell thought they some day would be—"chary of deserting the firm ground of measurable phenomena for excursions into the subjective" And plenty of those who insist the role of hypothetical individual expectations and preferences in economic analysis is a basic one, themselves engage extensively in the work of analyzing objective economic measurements.

I have used the statistical consumption function as an example of a behavioristic equation. But the class of aggregative equations it illustrates is not yet very large. There are, of course, a considerable number of current input equations in the Leontief model. Also there are the import function (a kind of step-child of the consumption function); the personal income size distribution pattern (the equation formerly used to state this pattern was called Pareto's Law); and the secular pattern in the functional distribution of national income. There are few others.

These not too stable behavior patterns may be contrasted with an-

other much more stable type of equation that I propose to call conceptual. Conceptual equations do not contain parameters that have been determined as statistical best fits as do behavioristic equations; and most such equations can be applied to future periods with great confidence.

More economists have disparaged conceptual equations than have disparaged the consumption function. It is because they have used the designation "definitional" in this connection to imply tautology that I suggest the substitute word, conceptual. Despite disparagement, conceptual equations are today the mainstay of quantitative economic analysis. And despite the implication of tautology a number of them are not quite true,—each of these imperfect equations contains an error term much as does the consumption function There are three broad classes of conceptual equations: those that assert that a whole equals the sum of its parts; those that assert that product equals multiplicand times multiplier, and those that assert a balance of debits and credits.

The second kind of conceptual equation is used in analyzing a dollar volume time series into a price-index series and a physical volume series. This form of factoring analysis was somewhat widely used in the early 1920's; it was also somewhat widely and seriously abused With the vast increase in available data and with a better dissemination of the necessary know-how, the use of this kind of quantitative analysis has been greatly extended during the past thirty years, and it has in general been tightened up by far more attention to detail. Abuses have not disappeared, but the grosser ones are now quite generally recognized as such.

Before taking up the other kinds of conceptual equations, let me call attention to an issue—or rather a group of issues—that the increased use of statistics has injected into economics If these issues can be summed up in a single question it is this, "Which should the quantitative economic analyst emphasize, his models or the complexity and ever changing nature of the world he is investigating?" One of the issues that stems from this question has already been noted, i.e , How fully should the quantitative analyst understand the economic measurements with which he works? Another and deeper one is, Should he devote his efforts to developing and improving a comprehensive model—say one along Keynesian lines, or a Leontief model? Alternatively, should he assume that such a schematic approach would be likely to impose unwise constraints on his inquiries? Or should he take some middle ground? I shall not stop to list other issues that stem

from the basic question of what to emphasize in quantitative analysis. Suffice it to say that together they divide the quantitative workers, not into two distinct camps, but rather into something like a spectrum.

Equations that assert the whole equals the sum of its parts make possible an obvious but basic type of economic analysis Thus one who wants to understand the behavior of a total like the Federal Reserve index of industrial production may wish to resolve it into its various industry components, and may, to bring out some influence at work during a particular period, focus his attention on a specially designed grouping of the components that are sensitive to this influence Economists who emphasize the complexity and changing nature of their subject are likely to rely heavily on this kind of quantitative analysis, and to feel that preoccupation with a comprehensive model might be a serious handicap when it comes to discovering significant special purpose groupings of components Equally, those who emphasize models may hold that preoccupation with *ad hoc* component analyses and a detailed historical approach is prejudicial to the discovery of stable economic behavior patterns Probably both groups are right.

Aggregative debit-credit equations are social accounting equations Most of the development of this type of equation—and it has been a substantial one—has come in the last thirty years Hardly anybody has bothered to call either component equations or factor equations tautological This charge has been directed against the social accounting equations. And the fact that they have been singled out in this way seems to be a tacit recognition of their high theoretical importance.

There are two reasons why a social accounting equation should not be considered a definitional equation First, every variable in a social accounting equation is either actually or potentially directly measurable, and thus has or can have an operational definition that is independent of the equation.[9] Second, a social accounting equation is significant theoretically because it expresses a significant economic adjustment. Thus the gross savings and investment account (gross S = gross I), against which the charge of tautology has most frequently been brought, reflects the equilibrating of supply and demand in the loan and security markets, although in the form in which it appears in the Department of Commerce national income and product accounts the supply of and demand for funds are not brought out very clearly.[10]

[9] Admittedly when a social accounting equation is used to provide a residual estimate of a variable, it serves *pro tem* as a definitional equation But this is not its main purpose, indeed its analytical usefulness may well be increased when a direct estimate or measure of the variable is developed

[10] The writer has discussed this point more fully in *A Study of Moneyflows in the United States*, New York National Bureau of Economic Research 1952, especially pp 246–60

Anyhow it should be obvious that this equation is not merely true by definition; the error term in the 1948 account for the United States was 5 per cent of gross investment.

Because social accounting equations have substantial theoretical significance, the recent rapid growth of social accounting has exerted an important influence on the development of economic thought. Let us consider briefly how.

For one thing, it has helped to correct a formerly somewhat prevalent misconception. Before we had much in the way of social accounting measurements it was frequently supposed that aggregate demand might be less than or greater than gross national product This misconception took various forms, but I shall cite only one of them, a theory of the cycle that had many adherents in the late '20's Its authors, Messrs Foster and Catchings, stated it in terms of what they called "the annual equation" or "balance of output and demand". This is clearly a social accounting equation. But they held, "The year is the shortest period of time within which we may reasonably hope to approach closely to a balance". And "The annual equation may be upset As a matter of fact, every recession in business activity is marked by this kind of overproduction or by the fear that it is imminent".[11] With the growth of national income and product statistics and of the practice of assigning these measurements a central place in aggregative analysis, economists have become cautious about suggesting any imbalance in a social account except a statistical discrepancy. Causal hypotheses involving a social accounting imbalance have been largely replaced by causal hypotheses depending on an imbalance in the subjective realm of hypothetical individual plans and expectations

But the really important influence of social accounting on economic thought has been the constructive one Before we had much in the way of social accounting measurements, aggregative analysis was rather like a ship without chart and compass Current business analysis it was called at the time. It lacked a central core of basic concepts around which inquiries could be organized The available current time series were appraised as business indicators or business barometers, the primary considerations in the appraisal were somewhat mechanically determined properties—cyclical lags and leads and cyclical sensitivity. Various selections of series with approved properties were combined according to various more or less arbitrary formulas into indexes of business activity During the last 20 years the system of social accounts has come very largely to replace these indexes Today a major consider-

[11] Foster, William Trufant and Catchings, Waddill, *Profits*, New York Houghton Mifflin Company, 1925, 249–50

ation in appraising a time series is its relation to this system; the GNP account and the sector or demand accounts that interlock with it. The system portrays the economic circuit which, at least since the time of Quesnay and Smith, had been recognized as playing a central role in allocating resources and determining the composition and distribution of product in the more industrialized countries The accounts have added precision to our understanding of the circuit, and they have given us measures of the main component flows of which it consists. Some time series report component flows, and are appraised on the basis of their roles in the circuit. Others, like prices and interest rates, are appraised in terms of the way they influence the circuit flow. Still others, like employment and unemployment, are regarded as resulting from the operation of the circuit. Thus the social accounts have helped to give aggregative economic analysis a sense of direction and a balanced perspective.

Note that I say, "helped"; they have not done this all alone A statement of an accounting balance is not a mere truism, but it is a rather colorless assertion, even when it asserts a balance of supply and demand It reports a correlation, it does not, by itself, indicate causation. Aggregative analysis has more of a sense of direction—and a sharper focus, too—than mere statements of accounting balance could impart. This improved direction and focus are to an important extent the results of a wide acceptance of a causal hypothesis proposed by both Keynes and the Stockholm school—the hypothesis that changes in the level of GNP are brought about mainly by changes in aggregate demand. Thus time series that report components of aggregate demand have a central place in aggregative quantitative analysis, and interpretations of the government account, the S and I account, the rest of the world account, and the personal account are directed toward explaining the behavior of the four major components of aggregate demand—government demand, private domestic investment, net foreign demand, and personal consumption expenditure. Without the social accounts there could be no such interpretations; without the hypothesis of the general primacy of aggregate demand the interpretations would not focus in this way.

Mitchell's forecast suggested that "the men now entering upon careers of research may go far toward establishing economics as a quantitative science". Despite the lingering defects in our statistical specifications and despite the persistence of subjective rationalizations of economic behavior there can be little doubt that the advance in quantitative aggregative analysis that has already taken place repre-

sents a long step toward the development of a quantitative economic science. And the period of the forecast still has twelve or fifteen years to run.

Progress has been substantial; but it has also been lopsided. Superficially, at least, micro-economic analysis seems to have been largely immune to the transforming influence of statistics, and empirical explorations of this phase of theory have not been very numerous. It is true something has been done toward developing statistical demand and cost curves; but there was rather more interest in statistical demand curves thirty years ago than there is today And the most significant development in micro-economic model analysis that has occurred in the past-thirty-odd years, that relating to monopolistic competition, has not to date proven very amenable to statistical exploration.

It is tempting to suggest a kind of cumulative disequilibrium theory to explain this lopsidedness The progress of quantitative empirical analysis in the macro-economic field seemed to offer a prospect of more progress Consequently the field attracted those economists who had an aptitude for quantitative analysis, and they neglected the cultivation of micro-economics. To the extent that such a theory has merit—and it seems to fit the declining interest in statistical demand curves—the lopsidedness is a defect that time should help to cure,—perhaps is already in process of curing.

But this is probably only a partial explanation of the lopsidedness. The push to divide economic theory along the macro-micro line came from persons interested in macro-economics The division facilitated work in macro-economics; it was in some ways awkward for the micro-economist. His concern is presumably with the various parts of the economy rather than the whole But most of these parts had long before been assigned to various special fields, e g , labor problems, agriculture, transportation and public utilities, private finance, industrial combination and competition And the effective application of quantitative analysis to a particular micro-economic problem is likely to call for detailed familiarity with one of these special fields. The bulk of the statistical work on micro-economic problems that has been done during the last three decades has been done by workers in the special fields. Certainly the result has been a substantial improvement in our understanding of the several parts of the economy, but the accomplishment here is a piecemeal affair that consists of a multitude of scattered bits of new knowledge. The cobweb theorem and the Federal Reserve elements analysis are striking, though hardly typical examples. It would be difficult to give a reasonable summary characterization of

what has been accomplished and I shall not attempt one. But it seems fair to say that the many scattered bits do not yet add up to a radical change comparable to that in macro-economics.

These comments, so far as they go, attribute the contrast between macro-economic and micro-economic quantitative analysis to influences that derive from the existing division of academic labor. But I suspect the contrast is due partly, too, to subject matter In the macro-economic field a problem has been singled out that is peculiarly amenable to statistical investigation, the problem of quarter-to-quarter and year-to-year changes in aggregate demand. The question is, "How do you account for the cyclical behavior of this aggregate quantity?" And a very large part of the answer can apparently be given in terms of the behavior of other economic variables, i e , in quantitative terms A very large part of the answer, but not all of it Qualitative factors, such as technological or legal changes, often have to be taken into account Still I think the fact that workers could afford to rely heavily on quantitative analysis in investigating business cycles has been a highly significant contributing circumstance in the progress this type of analysis has made And since the progress has been lopsided, it is pertinent that the circumstances are lopsided, too In the micro-economic field qualitative factors loom much larger. Witness the amount of time a specialist in labor, in railroads, in banking, or industrial combination and competition devotes to a study of the law as it impinges on his field If the progress of micro-quantitative analysis has not been very spectacular, it is partly because many of the problems of the special fields of economics do not lend themselves to a predominantly quantitative approach

Mitchell's address was primarily a forecast of the development of objective quantitative analysis But he took care to emphasize the complementary relationship between qualitative institutional analysis and quantitative analysis and the bearing of both on questions of policy and economic welfare He said in part, "quantitative work cannot dispense with distinctions of quality. . Indeed qualitative work itself will gain in power, scope and interest as we make use of . . . more reliable measurements" Out of "quantitative economics we may expect to come a close scrutiny of our pecuniary institutions and our efficiency in producing and distributing goods . Economists will concentrate . . to an increasing degree upon economic institutions . . Quantitative analysis promises . . to increase the range of objective criteria by which we judge economic welfare".

This portion of his forecast, to date at least, seems to have been

wrong. During the first third of the twentieth century economic theorists did a large amount of valuable work on the qualitative analysis of pecuniary and other economic institutions and their significance for economic welfare, but there has been very little of this type of inquiry since the appearance of Keynes's *General Theory*. True, a good deal has been accomplished in detail in the special fields, but statistics do not appear to have contributed any special impetus to the accomplishment. And in the field of theory they have, if anything, helped to divert attention from qualitative, institutional inquiries

Recently, however, the outlook for this type of work has brightened We now have a new division of academic-economic labor An increasing number of economists are concentrating their efforts on the problems of economic development and on comparing the institutional structures of different economies In this area of research the complementary relationship between qualitative, institutional analysis and quantitative analysis is particularly close It is too early yet to appraise the impact of statistical measurements here for we are only beginning to assemble a stock of this kind of information, and the field is still very new But its problems clearly call for a balanced combination of qualitative and quantitative analysis and this combination may some day have important repercussions on general economic theory Quite possibly there are to be found here the makings of a delayed response to the statistical stimulus somewhat along the lines Mitchell visualized [12] Certainly he was right when he said "qualitative analysis . . cannot be dispensed with "

[12] Cf comment by Domar, Evsey D , "Methodological developments" in *A Survey of Contemporary Economics*, Vol II, New York 1952, 455

FEDERAL TRADE COMMISSION REPORT ON CHANGES IN CONCENTRATION IN MANUFACTURING*

M A ADELMAN
Massachusetts Institute of Technology

STATISTICS relating to the number, size distribution, and other characteristics of the four-million-odd business firms in the United States are of obvious importance for public policy and the advancement of economic knowledge. The Federal Trade Commission is in part an original source of basic statistics, in larger part a "primary processor" of raw data supplied by others. In discharging both functions it has issued studies of concentration in the American economy.

The last previous study was characterized in this *Journal* as "this unimpressive monograph."[1] On the average, it is doubtful that anything better can be said for their latest effort; some of it is novel and useful, some of it is neither, and some (the most publicized portion) is egregiously bad More important than any over-all judgment, however, is a detailed examination.

The Report deals with changes in concentration in individual (four-digit) industries, and concentration in "manufacturing as a whole" For individual industries, the Report reprints the concentration ratios for 1947 compiled by the Census Bureau and published by the Celler Committee in 1949 For one third of all industries a comparison was possible with 1935. The contribution of the Report in this area consists of a chapter (pp. 22–9) of running comment, too brief to be useful, on the statistics There is no attempt at a measure of central tendency.[2]

Changes in concentration are related to changes in other variables. "[G]enerally speaking, the larger the expansion in number of plants in an industry, the more likely was a decline in concentration" (p 51); and concentration also tended to decline in industries with rapidly increasing production (54–7). Increase in concentration of physical establishments (plants) was accompanied by increase of concentration of companies (57–61) The information is presented in scatter diagrams;

* *Report of the Federal Trade Commission on Concentration in Manufacturing*, Washington Government Printing Office, 1954, Pp 153 $0 45.

[1] Stigler, George J., book review, *Journal of the American Statistical Association*, 46 (1951), 403–5. see also Kaysen, Carl, book review, *Review of Economics and Statistics*, 36 (1954), 107–9

[2] Gideon Rosenbluth found that the weighted average concentration ratio, corrected for a certain mathematical bias, declined by two percentage points during 1935–47, owing to the faster growth of the less concentrated industries On the basis of a stabilized population, that is, there would have been no decrease or increase See "Measures of Concentration," *Proceedings of the Conference on Business Concentration and Price Policy*, Princeton Princeton University Press in 1955, for the National Bureau of Economic Research, Pp. 79–83.

to this reviewer's eye, the association appears significant in each case, but it would have been a considerable improvement to have computed correlation coefficients and their standard errors, and to have supplied more than the few and sometimes questionable explanations There follows a generally useful chapter discussing twelve industries which experienced unusually large increases or decreases in concentration; the quality of the discussion is uneven, and one would hope for a more discriminating use of source materials. Thus, it may (or may not) be true that "independent barrel makers were at a competitive disadvantage in obtaining steel during periods of shortage" (p 81). But the only support for this conclusion is that an interested party, the vice-president of one of these companies, complained "in a letter to Smaller Business of America, Inc " This kind of unverified complaint may at best be a lead, it is not evidence.

By far the best part of the Report is in certain appendixes Appendix B computes indicators of homogeneity, measured by an industry's shipments of its primary products as a per cent of its shipments of all products, and as a per cent of all industries' shipments of the primary products. About one-seventh of all industries were eliminated from comparison because they fell short of a homogeneity standard of 66.7 per cent for both indicators One would appreciate some explanation of why the particular cutoff point was chosen, but the procedure is certainly correct, and the tables are useful for reference. Appendix C covers the reliability of measures, compiled by the Census Bureau, of the concentration of employment in individual industries in 1950 These are based on the Annual Survey of Manufactures, and are therefore affected by sampling errors, including possible failure to include all the largest concerns. The higher the concentration ratio, the wider the range corresponding to a given confidence interval This was to be expected from the general nature of the reliance on the largest units of a highly skewed distribution as a means of achieving a more efficient sample design. The resulting error of bias (in a non-probability survey) or lesser degree of precision (in a probability survey) are most likely during periods of rapid change Also of importance for short-period changes is a special tabulation by Census showing what changes would occur if certain large establishments were classified in industries representing their primary products in 1950 rather than in 1947, to which the Annual Survey adheres This "resistance to reclassification" is adopted by the Census Bureau because of the anomalies that might result when the products of a large establishment in one year were (say) 55 per cent in industry A and 45 per cent in industry B. In a

succeeding year the percentages might be reversed. Unless some resistance principle were adopted, the result would be a spurious decline in the output of industry A and a spurious increase in the output of B. The Report does not give any coherent explanation of this practice and its justification, but it does classify 247 industries by concentration ratios, standard error, and allowance for resistance effect The reviewer would consider these as more than mere detail improvements greater practical use and also improvements in concept are both to be found along this line.

Much interest will attach to the Report's computation of concentration in "manufacturing as a whole" This consists of the aggregate sales-plus-interplant transfers of the 200 largest manufacturing corporations as a per cent of sales-plus-interplant transfers of all manufacturing establishments There is no industrial breakdown of the 200. This measure of concentration was 37.7 per cent in 1935, and 40.5 per cent in 1950.

The use of sales-plus-interplant transfers as the dimension of size and concentration was not necessary for a 1935–1950 comparison. The 1935 source[3] tabulated the largest 200 manufacturers as accounting for 31.1 per cent of value added by manufacture, for 24 8 per cent of employment, and for 37.7 per cent of value of product (sales). Why did the Report use this measure of size? Value of shipments to measure concentration in individual industries may be defensible as a practical matter or for certain purposes even in theory. But there is no such justification for using sales as a measure of size when grouping corporations together across industry lines. The result is massive and uncontrolled double counting It is almost embarrassing to have to point out that the size of a business firm is defined by its own employees assets, or contribution to national product—and not somebody else's To lump together, on the basis of their sales, corporations with widely varying ratios of sales to employment, assets, or value-added, is to aggregate horses and apples

Furthermore, this worst of measures for any given year will make it impossible to compare the 1950 estimates with those for future years although the opening sentence of the Report states this as a primary aim (p. 1) Value of shipments for all manufacturing has not been published since 1939. The Report never states this fact, only referring (p 106) to "the Federal Trade Commission's estimate of the value of shipments of all manufacturing industries" On the next page the

[3] National Resources Committee, *The Structure of the American Economy* (Washington Government Printing Office, 1939), Appendix 9.

distinct impression is given, no doubt inadvertently, that this nonexistent statistic can even be assigned a standard error! In any event, had the present Census rule existed in 1935, the comparison made in this Report would have been impossible. Unless the Bureau changes its practice, it will be impossible to compare this estimate with any future estimates.

Moreover, the use of value of shipments involves much unnecessary labor. For computing the share of the largest 200, and for no other purpose, the Report uses the item of interplant transfers and adds it to sales No less than 35 per cent of the 1950 universe figure (sales-plus-interplant transfers of all manufacturing establishments) had to be estimated, with the usual approximations and assumptions, some better than others. All this adjustment would have been avoided simply by having the largest 200 respondents supply value-added (and/or employment) instead of interplant transfers Then published universe figures would have been available, the work would have been less, and there would have been two valid measures of concentration

The defects of value of product are compounded in measuring *changes* in size and concentration over any extended time period With the exception of assets, which are essentially a moving average, centered at a point before the date of record, all measures of concentration based on one-year periods and current values are distorted somewhat by price changes and by the business cycle As to price changes, value of products is particularly bad because it reflects fluctuations not only inside but also outside the firm, the industry, and all manufacturing. Thus, if the price of raw cotton were to rise more than the price of raw materials generally, the value of product of cotton textile mills and apparel manufactures would *ceteris paribus* rise more rapidly than the value of product of all manufacturing. Since cotton textiles and apparel are a relatively small-business industry, there would be a spurious decrease of concentration in "manufacturing as a whole" We need only recall that in 1935–1950 the wholesale price index doubled, with many divergences among industrial groups

As for cyclical changes in 1935 manufacturing capacity was roughly about half utilized[4]—which means that many industries and companies, mostly in durable goods, were spuriously "small." The year 1950, on the other hand, was one of full or over-full employment Its second six months saw a speculative Korean war boom, which temporarily over-stimulated the real output of some industries (It also

[4] The Federal Reserve Board index of manufacturing production (1935–39=100) stood at 87 in 1935, and at 168 in 1941, durable goods production was respectively 83 and 201.

caused speculative price movements, all irrelevant but all recorded in the 1950 value of shipments) An examination of 1935–50 corporate sales by industry groups shows that the combined effect of price changes and cyclical changes was to increase the sales of the big-business industries, who account for practically all the sales of the top 200, substantially more than manufacturing corporate sales generally. This probably more than accounts for any "increase" in the FTC measure. (*Statistics of Income* shows that in 1931 the 139 largest manufacturing corporations held 46 5 per cent of the assets of all manufacturing corporations: in 1949 the largest 142 held only 43 4 per cent.) But whether concentration increased or decreased over this period, a measure based on sales-plus-interplant shipments deserves no credence; the figure has no meaning.

Obviously, this report was prepared before a majority of the present Commission assumed their posts. Of the new arrivals, the Chairman has expressed strong support of more and better economic and statistical research by the agency. This report contains two excellent guideposts. The broad-brush portrayal of increasing concentration which drew attention in the press and the Congress is based on a technique which was outmoded at least a quarter-century ago, when the pioneering estimates of Gardiner C. Means appeared; the results mean nothing. In contrast, the undramatic appendixes which get down to the actual job of measuring concentration are a genuine contribution and give promise of more.

COLONIAL SOCIAL ACCOUNTING*

WILLIAM O. JONES
Stanford University

PHYLLIS DEANE's report on the experiment with national income accounting applied to colonial territories, published in 1948 as an *Occasional Paper* [2] of the National Institute of Economic and Social Research, concluded with the statement that "a more comprehensive and direct knowledge of the social and economic structure of Central African peoples was essential if a satisfactory framework was to be evolved". In order to obtain this knowledge the investigation was to be transferred to the African colonies themselves, and direct field enquiry used to supplement the existing information *Colonial Social Accounting* is a final report on that enquiry It contains revised estimates of the national incomes of Northern Rhodesia and Nyasaland (Parts II and III) together with an account of village economic surveys and of some of the conceptual and operational difficulties involved (Part IV), and detailed summary of the sources of the estimates (Appendixes I and II).

The most valuable chapters are those describing the problems that must be solved if native African economies are to be forced into this mold that fits European and American economies only imperfectly. Unfortunately these sections will be read and quoted less widely than the estimates themselves The estimates of per capita income of the African population are already being quoted as if they were reliable measures of level of living, and will undoubtedly attain wide currency. It might have been better to arrange the parts so that the qualifications precede, rather than follow the estimates It is all too easy now to read the first 100 pages, containing all of the income tables with little suggestion of their fundamental untrustworthiness, and stop. But perhaps it was reasoned that people rarely pay any attention to the qualifications placed on statistical tables.

The book itself speaks with two voices: the straight face with which the estimates are presented is disturbingly inconsistent with the bewilderment expressed in later chapters over the problem of evaluating native activities in units commensurable with those used for the European part of the economy. Deane has provided a handsome target for the reader critical of attempts to assign money values to goods and

* A review article on Deane, Phyllis, *Colonial Social Accounting*. Cambridge, England The National Institute of Economic and Social Research, Economic and Social Studies XI, 1953 Pp. xv, 360. $10.00.

services that are rarely exchanged for money and to the multifarious activities having no value at all in American or European eyes; she has also provided some of the rocks to throw, and has even tossed a few herself. Indeed, it is not altogether clear whether she feels the experiment in "colonial social accounting" to have been a success or a failure The concluding paragraph of the book begins with the following disheartening statement (p 228):

"These, however, are the basic truths about colonial social accounting, and it is difficult to see how they can honestly be thrust into the background The statistical material is inadequate for purposes of precise and intelligent analysis, and the concepts which are applicable to a money-exchange economy mean relatively little in the context of a subsistence economy . "

Were it not for the word "precise" in the foregoing, this would seem to be as full an admission of failure as the most severe critic could ask. It is in sharp contrast with the cold tables of Parts II and III of the book in which the author records and combines estimates of incomes of Africans and Europeans without apparent embarrassment.

The experiment was of course not a failure It produced estimates of various components of the national incomes of the two territories, some of fair accuracy And even the total figures are probably good guesses as to the magnitudes they attempt to consider These are certainly estimates of national income in the restricted sense in which this term is customarily used. But it is seriously open to question whether the figures for the native sector of the economy may be taken either as indicators of level of living or of economic output They are estimates, valued in European measures, of those parts of native production that have counterparts in the European sector

Even the making of such estimates as these was a formidable task; the author performs the additional service of pointing out their many weaknesses and the wide ranges in which the true values seem to lie.

The problem Deane set herself was to assign a monetary value to the product or service of every "economic" activity Prices to be attached to each activity were derived from its contact, however slight, with the exchange economy. If 5/10ths of output was consumed by producers, 4/10ths bartered, and 1/10th sold, the part sold determined price This might be a very narrow market; it might also be an abnormal market, resorted to only in times of great need or affluence, or to obtain money for trivial or unusual purchases [1]

[1] There is in fact no satisfactory way of obtaining the retail prices prevailing in native markets. Each sale is usually at a price reached only after bargaining, and the price varies from market to market and even from buyer to buyer Prices quoted to Europeans may be higher or lower than those at which sales are made to Africans.

A simpler solution might have been adopted, that of ignoring all transactions except those in which money is involved The Central African Statistical Office estimates Net National Income for Northern Rhodesia generated in the money economy only, explaining that "it was felt advisable to omit any statement of the value of subsistence output as it could only be a notional figure that could not be checked or corrected in any way . " and would probably not be additive to the estimate for the money economy [1].

But ignoring the "subsistence economy" in Northern Rhodesia and Nyasaland is a pretty serious business for one who, like Deane, believes that "the need is for a design of accounts which shows up the distribution of the national income among the different groups, the nature of each one's contribution to it, and the characteristic composition of each group's standard [level] of living" (p 5) In Central Africa, most of the native population satisfies its economic wants in the subsistence sector, and failure to include that part of national output leads to a very peculiar measure of individual income

When account is taken of the statistical problems involved in obtaining the kind of measure Deane sought, and the economic problems involved in interpreting the measure once it was obtained, the solution adopted by the Central African Statistical Office appears to have merit. National income generated in the money economy is an understandable magnitude It does not pretend to tell anything about economic welfare of the entire society; it measures change in comparatively simple terms, and it summarizes information about that part of the economy in which the consequences of governmental intervention are most nearly predictable

Administrative budgets in the colonies have tended to be tight, with little money available to support even the most elementary statistical services. There has never been a census of population, of agriculture, or of business in the colonies considered by Deane Estimates of population and agriculture are usually built up from reports of District Commissioners on tour. The population estimates made from "tour counts" are very rough indeed and during the years to which the study applies there had been very little touring [2]

Estimates of agricultural production are much worse. To find a value for native agricultural output to put in the Nyasaland table (it is the largest individual value in that table), the following assumptions were involved (pp 317–18).

---

[2] A sample census of population in Northern Rhodesia was taken in 1950, too late for use in the study But Deane believes a sample census little more accurate than District Commissioners' counts (p 231). Population estimates from the sample census vary little from those used by Deane.

1. Assume "that the average cultivated acreage per family in Southern Province was 2 acres and in Northern and Central Provinces 3 acres, . . . "

2. Assume "that the number of families was equivalent to the number of married or widowed women as shown in the 1945 Census, . . . " Of this census the Colonial Office says [3] "it claimed to be no more than a useful, and in the aggregate a fairly accurate estimate of the African population based on a count".

3. Assume "that maize or its value equivalent was grown on about 75 per cent, groundnuts, beans or their value equivalent on 20 per cent, and cassava or its value equivalent on 5 per cent"

4. Assume yields per acre for each of these three groups of crops.

5 Assume prices for each of these three groups of crops.

6. Assume "that each individual consumed in a year an average of 10 lb. of fruit, green leaves, etc., not grown on the cultivated acreage, . . . "

7. Assume that the fruit and green leaves can be valued at 2 pounds for a penny.

By procedures such as this Deane arrived at her estimate that African income from independent agriculture amounted to £3,561,000 in 1945 (p. 98).

Even when it comes to estimates for the European part of the population, the author confesses some uneasiness. In Northern Rhodesia the number of Europeans earning incomes was only 8,225 (the number in Nyasaland was 1,200), all incomes over £700 were subject to tax and the 1946 Census of non-Africans contained detailed questions about incomes, and three industries, mines, railways, and government, employed five-eighths of all Europeans with income, yet Deane states that the margin of error in the estimate of total money incomes of European individuals was about 6 per cent (pp. 237–243). The error in the estimate of total money earnings of Asians and Colored (mulattos) must have been larger.

But Europeans, Asians, and Colored make up a very small part of the total population (about 1 per cent in Northern Rhodesia, about one-quarter of 1 per cent in Nyasaland), and anyone who has read Deane's account of how estimates of incomes for this part of the population were made, of the cross-checking, comparing, and calls-back, must be convinced that the errors here are not so large as to be damaging. It is in the estimates for the other 99 per cent of the population, the Africans, who have never been counted, but who are the population

of the country, in any normal meaning of the word, that the trouble lies.

We can divide the income, or the output, or the expenditures of the African population into two main parts: that which is in and a part of the European-Asian economic order, and that which is completely within the old native order. Probably no individual will be completely part of either order, nor will many activities. The native houseboy who is paid in money, food, and lodging, but whose food is African food and whose lodging may be African lodging, both quite different from those favored by his employer, and who still preserves ties with the old order through his village and through a town social system that has close links with the past, clearly produces and receives income in both worlds On the other hand, the farming family that produces almost all of its own food, and makes its own house and furniture, will receive a little money income from the sale of farm products or from a family member employed by Europeans, and will buy a little—pots and pans, cloth, salt, soap for example—in the markets.

It is probably the output and income of the "non-money" part of the economy that is most important to the native population, and it is in the valuation of this income that the investigator runs into serious trouble The trouble takes two forms; one can be solved conceptually, the other so far has not been The first problem is that of counting or measuring the various products and services in units peculiar to each—in number of pounds or gallons or yards, in number of chairs or stools or tables or houses, in hours or minutes during which the service is enjoyed, or even in the decreases in numbers of stillborn, in deaths at childbirth or from particular diseases The second problem is one of the oldest known to economists, how to convert all these disparate counts and measures into one, and that one somehow to be a measure of value, of total economic output, or of welfare.

Deane has labored courageously at the first task. She has probably done as well as could be done. (I can not help wishing she had shown us a little more of the way in which she arrived at the output of native agriculture) In her attempt to measure the magnitude of native output she proceeded with great care and understanding The chapters in Part IV that describe the job of trying to get aggregate measures for a fragmented, largely illiterate society, self-contained to a remarkable degree, and with little money should be read by all who contemplate such an undertaking or who wish to use her estimates These chapters also contain a great deal of useful and rare information on money income, property, purchases, and gifts in the communities studied.

The second problem, however, she really does not try to solve at all Instead she uses a device similar to that employed by the Central African Statistical Office, only going a little further. The total tonnage of corn produced is multiplied by the price for which some corn was sold in some native markets; the housing supplied to miners as a part of wages is translated into money at the estimated average rent paid for rooms and dwellings actually rented for cash (p. 247), plates, cups, and knives were valued at "the average price for all the recorded transactions in that commodity for the area concerned" (p. 193). By such means as these it is possible to translate a wide variety of measures into money What the meaning of the translated values is, if they have any at all, is another matter.

Deane points out clearly one of the difficulties in such translations when she describes her handling in the village surveys of "Transactions Without Cash" (p 195).

"It seems, for example, that barter transactions in goods and services for immediate consumption are both common and casual A man may barter a chick for some beer because he is temporarily short of cash and a beer party is in progress. A woman may barter a few groundnuts for a cupful of salt because she has not time to go to the store for it A man may prescribe for another's cough and receive a pumpkin in exchange. All these are small and ephemeral transactions, often entirely unconnected with current exchange values, . "

In Chapter XIV there is recognition of the importance of family production together with a valuable summary statement of the problem of estimating it, and of the artificiality of the technique used to include it in the estimates

How important native agriculture is to the total economy can only be suggested by the estimates, for 1945 it contributes 40 per cent of national income for Nyasaland (p. 98), and 23 per cent for Northern Rhodesia (p 64), where the mining industry accounts for a slightly larger share. It is difficult to appraise the possible error in the estimate of native agricultural output, even accepting the author's device for translating into money. That the true values may be twice as great as those she uses is not at all unlikely.

Estimation of the national income of dual economies like those considered here involves essentially the same sort of problem as is encountered in comparison of national incomes of two separate economies. The difference is purely arithmetic; in one case interest attaches to a sum, in the other to a difference.

But the nature of the difficulty is somewhat obscured when one and

the same currency is used in both parts of the economy, as it is in the native and European sectors of the colonies. If different currencies were used, attention would at once be directed to finding an appropriate relationship between them, and it could be hoped that the pitfalls in such comparison would become apparent

Economists have achieved some success in comparing incomes of economies that are similar in their organization and in their values, but even then the best we can hope for is that a high percentage of the items in the two economies correspond and are similarly valued Complete correspondence in the nature and the valuation of all economic goods and services is too much to expect, and we must always resign ourselves to a degree of incommensurability.

When two economies are as different as those of the native and European parts of colonial territories, the extent of correspondence of items and values is apt to be very small If, in addition, the money values of goods and services exchanged between the two economies are determined almost entirely by one of them, comparison of outputs in monetary terms will be misleading in the extreme.

I think it is not open to argument that the standards of values, or the standards of living in the strict sense, of the traditional native societies of Africa differ about as widely as is possible from the standard of the intrusive European societies As a result, goods and services produced by the native economy are apt to be improperly valued, or even completely overlooked, by the investigator who uses European standards to measure them.

The native standard of living may contain things that are not in the European standard at all, or if present are lost to the view of the statisticians Included here are the apparently pervasive desire for identification with and absorption in the village, tribe, or extended family. This may be valued for the security it provides in terms of food and shelter, or for the deeper security of belonging, of being a part of a greater whole in some emotional or psychological sense Perhaps leisure should also be included here, although leisure activities are valued in the European world I should be very much inclined to include as one part of the standard a way of life that is not bound by hours of the clock, that permits flexibility in the nature of activities and in the time devoted to them.

Other elements in the native standard may have their counterparts in the European world, but carry much lower values Native religious or magical ceremonies, rites accompanying birth, puberty, marriage, and death, the satisfaction of participating in cooperative undertak-

ings, the pleasures derived from taking part in the activities of the market all have their parallels in other societies. They do not, however, appear in national income accounts except as they involve expenditures, such as payment to build a church, initiation fees for a lodge, the purchase of confirmation gowns, or the cost of attending a national convention.

There are also many components of the native standard not assigned a value by Europeans even though valued European counterparts exist. Included here are dances, beer parties, and other entertainments, cooperative services and labor exchange, works of art, and sport, not only for itself alone, but sport inextricably interwoven with provision of material needs, such as fishing and hunting, and even some farming activities, as for example the cutting of high tree branches in new clearings.

Beyond the problems introduced by the considerable lack of correspondence between elements of the two economies, additional complications may arise because of the peculiar role of money in the native economy. Deane's comments already quoted on relative values of bartered goods may also apply to merchandise offered for sale. Money transactions may be completely trivial to the native living in the village, either because the need for money is small or because the share of total output sold is small On the other hand, the price may be trivial because the desire either for money or for goods is urgent but at the same time sporadic.

Moreover, if the use of money prices is unfamiliar, and if rates of exchange of goods and services tend to be determined by tradition and hence inflexible, equivalences that might be inferred from trade are again quite misleading

It is a common observation that the native's desire for money may be satisfied fairly quickly, and there is frequent complaint of the way in which hired laborers stop work when they have obtained some desired, usually small sum This undoubtedly reflects the availability of goods in the market· a limited variety of useful merchandise that is within the native's horizon of possible acquisitions, other highly prized articles completely outside his potential income It may well be that so long as the native income remains low, the demand for money will likewise be low, but that after some higher income level is reached, demand for money will shoot up sharply.

Also associated with the general shortage of money may be an overvaluation of currency. If a very small supply of coins and bills is avail-

able to handle an increasing volume of trade within the native economy, there may be a continuing deflationary effect not corrected by an inflow through the narrow channel of trade connecting with the European economy.

These conjectures about the place of money in the native economy require further examination and comparison with the facts Taken with what is known about the great differences between native and European living standards, however, they go far to explain why investigations conducted along the lines followed by Deane give what appear to be nonsensical answers.

Incomes of £6 per head per year arrived at by Deane are in no way an appraisal of economic output, as valued by those who produce it, or of level of living, as valued by those who experience it. Valuation of housing at 6s. per employed miner per month is a similarly meaningless figure. It is not implied that these numbers are without meaning to those who know native standards and ways of life, but they simply cannot stand alone, sufficient in themselves, as measures to be added to the estimated value of European, Asian, and Colored incomes

Much of the foregoing may be interpreted as an attack on a straw man Certainly the reader is warned in the last half of the book of many of the dangers inherent in the figures But calculation of the magnitude called national income in itself implies meaning that cannot be supported by the facts.

If we expect the native society eventually to be remade in the image of the European, then, when that time comes, national income estimates of the kind made here will be just as valid as they are in Europe or the United States. In the meantime they may even serve as some sort of index of the extent to which the native economy has been Europeanized, although a better measure would seem to be that of national income originating in the money economy.

Colonial Social Accounting reports an honest attempt to apply to backward economies a technique that has had considerable success when applied to more clearly market-oriented economies Defects of measurement that are troublesome but probably not serious in America and Europe, such as those arising from appraisal of unpaid household labor, on the one hand, and from the differing values placed on leisure, fine cooking, spaciousness, or independence, on the other, here have proven sufficiently grave to jeopardize the entire enterprise.

For appraising level of living, for measuring changes in output and productivity, the estimates are unsatisfactory. Of course we can say

that output is only economic if Europeans think it so, and to the extent that they think it so, and let it go at that, but I find this a rather uninteresting solution It would be much better to recognize the incommensurability of the components and not aggregate them at all. Much more can be learned about output, productivity, and level of living by studies of agricultural yields of specific crops in pounds per acre and per man, by collection of data adequate to permit calculation of food balance sheets, by careful diet studies, by statistics of mortality, and by other similar investigations yielding results that have clear meaning in themselves. It would seem wiser for colonial statisticians to spend the limited funds available to them on studies such as the sample surveys of agriculture conducted in Nigeria in 1950–51 [4], and in Tanganyika in 1950 [7], on improvement of trade statistics [6], or even on detailed diet studies like those made in Nigeria by B M. Nicol [5]. Better trade statistics could greatly improve our knowledge of the money economy; improving agricultural statistics in some areas and remedying the total absence of such data in other areas would vastly improve our knowledge of native economic production; and even small samples of native diets, if taken carefully and reported in detail, would move us a long way forward in our knowledge of native welfare.

The priority job in colonial statistics should be to get better estimates of the important magnitudes. Even when this has been done, however, it may still be necessary to keep them separated, to resist the temptation to aggregate them Nor should we regard estimates of farm output and of food consumption as other than the crudest measures of welfare. More complete knowledge can come only as a result of extensive comparisons between native standards and levels of living. On the other hand the elementary measures can be used in comparing similar native economies and in measuring change over time Similarly, improved statistics for the money economy will provide more precise bases for forecasting the consequences of intervention in that area

If I criticize Deane's analysis, it is not because she has failed to recognize and discuss the problem of evaluating native product, but because she has then gone ahead in her calculations as if these problems did not exist Chapter IX, "Social Accounting for Primitive Communities", contains enough arguments *against* the use of national income accounting in primitive communities to stop all but the most enthusiastic devotees of the method. Yet however great the obstacles may be Deane forges resolutely ahead "It is almost always possible", she says, "to cite an instance where a money price has been charged within

recent memory and in the same general area On the other hand, it is probably safe to say that for the majority of economic activities there is no market value as it is understood in the money economies. When goods are not normally traded the prices of the goods that are traded do not reflect the value in resources used or in relative desirability of subsistence output" (p. 123).

She then cites as an example the sale of food during a period of shortage before harvest, properly noting this as an extreme case, but: "For all the most important commodities there is a price which has some kind of meaning, even if the meaning is irrelevant to the internal economy of the village" (p 123).

When it comes to the use of money within the native economy, Deane says, "although money is used freely and money value is a familiar concept in the semi-subsistence economy, the prices are not necessarily part of an integrated system. . . There is no common standard of value which would make it possible to consider each economic activity as a function of its drain on the community's total scarce resources or of its contribution to the community's total needs" (p. 124)

There is a great deal more in this chapter to the same effect. It provides effective argument against making and publishing the kind of tables that are the heart of Parts II and III. And yet the chapter concludes with the statement, "The best that can be said about a social accounting system when applied to a primitive economy, . . is that it takes all sectors and aspects of the national economy into account and that it provides a very rough and often one-sided indication of their relative contributions to total national economic activity" (p. 130).

There must be considerable doubt as to whether such accounts provide even a rough indication, there can be no doubt about their one-sidedness More consistent bias, as in the Central Statistical Office's estimates, would be preferable

REFERENCES

[1] Central African Office of Statistics, *The National Income and Social Accounts of Northern Rhodesia, 1945–1949,* Salisbury, Southern Rhodesia, May 1950, 1.

[2] Deane, Phyllis, *The Measurement of Colonial National Incomes An Experiment,* Cambridge, England National Institute of Economic and Social Research, Occasional Paper XII, 1948

[3] Great Britain Colonial Office, *An Economic Survey of the Colonial Territories 1951, Vol. I,* Col. No 281 (I), 1952. 36

[4] Nigeria, Department of Statistics, *Report on the Sample Census of Agriculture, 1950–51*, Lagos, 1953.
[5] Nicol, B. M., "The nutrition of Nigerian peasants with special reference to the effects of deficiencies of the vitamin B complex, vitamin A and animal protein," *British Journal of Nutrition*, 6 (1952), 34–55
[6] Seers, Dudley, "The role of national income estimates in the statistical policy of an underdeveloped area," *Review of Economic Studies*, XX (1952–53), 159–68.
[7] Tanganyika, *Report on the Sample Census of African Agriculture· 1950* (revised), prepared by the East African Statistical Department, November 1953.

CONCEPTS EMPLOYED IN LABOR FORCE MEASUREMENTS AND USES OF LABOR FORCE DATA

A. Ross Eckler, Gertrude Bancroft, and Robert Pearl
Bureau of the Census

Controversy over labor force concepts arises to a large extent because of the varied and often-conflicting uses of the data. The Census Bureau has tried to meet this problem by providing detailed information on major categories and by experimental work to measure borderline groups. Types of questions that have been examined are. difference between labor force participation during a year and a calendar week; underemployment of employed persons, strength of job attachments of persons not at work for various reasons, the appropriate classification of persons on the fringe of the labor force Efforts have been made also to reconcile statistics from households and from business and administrative records.

It is a well-accepted principle that concepts should be developed in the light of the uses to be made of the data being collected However, the uses of labor force data are manifold, and often conflicting. One of the important uses is to provide a comprehensive measure of changes in economic activity. There is general recognition that not only are the employment and unemployment totals important, but also that the changes in hours of work need to be watched, as well as changes in the number of those who are working on a part-time basis because of actions taken by plants to avoid lay-offs Changes in duration of unemployment are also used as an indication of the significance of a particular business readjustment.

Another important category of use, but one which varies markedly in public interest, is in connection with general manpower analysis At the time of large-scale defense effort or of actual mobilization, data on the labor-force status and other characteristics of the population are valuable in providing the basis for estimating manpower potentials under various assumptions.

A third important use is to give current information on some of the longer-run social and economic population changes which are more fully shown in the results of the decennial censuses The current labor force series reveals, for example, changes in the age of entrance into the labor force as trends toward the prolongation of schooling continue At the other end of the working life, they show the extent to which retirement is affected by social security and other provisions. Data on family composition, including the number of young children, show the

relationship between family duties and outside employment, and thus indicate the extent to which the proportion of women in the labor force continues its long-run rise

Other uses may be mentioned even more briefly. For example, the labor force series is the only comprehensive source of information on groups of workers to whom special attention is directed from time to time such as veterans, older workers, non-whites, domestic workers, the disabled, and others The data are of interest to some analysts as an approach to measurement of labor input, while others would like to relate them to the production of consumer income.

Not only do the uses exhibit wide variety, but there is also substantial difference of opinion as to the values to be attached to various features of the data At one and the same time, we are being urged to increase the accuracy of the figures, to expand the detail published, to speed up collection, to improve methodology but maintain consistency, to develop alternative measures, and to reconcile these data with other statistical series Those interested in the measurement of economic developments are likely to stress speed, accuracy, and sufficient continuity to make month-to-month and year-to-year comparisons significant. Others concerned with broad social and economic analysis do not require great accuracy and speed, but want a wide variety of information and many detailed cross classifications. Those studying a particular group would, of course, ask for the maximum amount of subdivision for that group Still others, interested in neither level of accuracy nor in consistency of classification, are ready to accept whatever is published and are impatient with technical sections describing sampling and non-sampling errors.

STATEMENT OF PRESENT CONCEPTS

The present labor-force concepts have been used for less than 20 years The 1940 Census gave up the "gainful worker" concept based upon the reporting of a gainful occupation and adopted in its place the concept of "activity or behavior" in a particular week During the period since 1940, the basic concepts have not been changed, but as will be described later, a good deal has been done in clarifying the concepts and in recognizing and attempting to remedy some of the deficiencies.

The present concepts have been described at so many places that a brief review should suffice at this time. The classification of a person by employment status is based primarily on his reported activity during a specific calendar week. If the person worked at all for pay or

profit during the week—or without pay for 15 hours or more in a family-operated enterprise—he is classified as working, and hence as employed, regardless of other activities. If he was not working but was looking for work, he is classified as unemployed. In order to obtain realistic current counts of the employed and unemployed, certain so-called "inactive" groups are added to each primary category. Included with the employed are persons who were neither working nor looking for work, but who had definite jobs or businesses from which they were temporarily absent the entire week for such reasons as vacation, illness, industrial dispute, and temporary (less than 30 day) layoff Included with the unemployed are those who would ordinarily have been looking for work but were not because of temporary illness, belief that no work was available in their line of work or in the community, or because they were awaiting recall to jobs from which they were on indefinite layoff. The civilian labor force includes the employed and the unemployed; the total labor force also includes the armed forces. Civilians neither employed nor unemployed are classified as not in the labor force. The basic classifications for the monthly series are applied only to the population 14 years old or over.

EXPERIMENTAL WORK FOR PROBLEM GROUPS

In the past 12 years much experimental work has been carried out under the guidance of interagency committees and subcommittees to explore a variety of problems in connection with the application of the labor force concepts described A brief account of them may be helpful as a basis for consideration of further steps that could be taken. Some of the research has pertained to what might be considered the "ground rules" covering such matters as age, time periods, etc , some with the collection of more detailed information on certain groups, and some with the collection of more information on the appreciable numbers of persons in certain marginal categories.

QUESTIONS RELATING TO AGE LIMIT AND PERIODS OF REFERENCE

Age. The exclusion of children under 14 years of age from the labor force statistics has been a matter of relatively little concern, except possibly to groups particularly interested in problems of child labor In fact, there has been some pressure to raise the age limit to 16 or 17 and to provide a cutoff at the other end of the scale, possibly at age 75, in order to eliminate some of the more marginal groups from the statistics

Nevertheless, substantial numbers of young children under 14 [illegible]ll

do farm work; and persons accustomed to finding their daily newspaper at the door might pay tribute to the economic contribution of nonfarm youngsters as well. On a few occasions, usually with outside sponsorship, the Bureau has collected labor-force statistics for children under 14 years of age.[1] In addition to their utility to groups interested in child-labor problems and possibly to labor input analysis, these data have been of some importance in the reconciliation of the Census Bureau's statistics with those based on employers' or establishment reports, which generally specify no age cutoff.

Use of Week as Time Reference. The use of a calendar week as the time reference for labor-force surveys has had generally widespread support. Measurements based on a week are not only a reflection of current status but are also less subject to the memory biases which arise in the use of a longer time period. When conditions are atypical during a specific week, however, and especially when that week represents the only measurement obtained in a given month, there are certain obvious disadvantages. Moreover, for some purposes, alternative types of measurements are desirable in order to obtain an adequate description of the complex labor market behavior of a large segment of the population.

The occurrence of a legal holiday during the survey week, of course, has a substantial impact on hours worked and will obscure month-to month movements in the workweek. Some experimental studies have been made to measure the effect of holidays on hours distributions and manhours worked and it is believed that rough adjustment factors could be developed through further research of this type.

Some of the limitations of the use of the calendar week approach are also met by a regular annual inquiry on work experience during the preceding calendar year. Persons are classified according to the number of weeks in which they worked during the year, whether they were working fulltime or part time while employed, and according to their principal occupational and industrial attachment. Although information of this type is somewhat out of date by the time it is compiled and contains deficiencies because of memory factors, it does in some ways provide a more comprehensive description of the adequacy of employment for the period covered than does an accumulation of the monthly statistics Moreover, for specific individuals, it may provide a more meaningful and typical classification.[2]

---

[1] For some of the results, see U S. Department of Labor, Bureau of Labor Statistics, *Caution Children Under 14 at Work*, January 24, 1951.

[2] For a summary of these surveys, see U. S. Bureau of the Census, *Current Population Reports*, Series P-50, Nos. 8, 15, 24, 35, 43, 48, and 54, the first three of these studies were based on a somewhat more limited approach than that described above.

A good illustration of the differences revealed by the two approaches may be found in an examination of the statistics for white and nonwhite women Based on the calendar week approach, the proportion of nonwhite women who are in the labor force has historically exceeded that of white women by a wide margin, suggesting more intensive economic activity on the part of the former. The annual surveys indicate, however, that the participation of nonwhite women is on a much more intermittent basis and that the annual per capita labor input of the two groups would show much less variation. In 1953, for example, only one-quarter of the nonwhite women with work experience had full-time work the year round; the comparable proportion for white women was 40 per cent. In any evaluation of these relationships, it should be borne in mind that for persons with intermittent employment the reports upon work during a preceding year are subject to considerable ranges of error

The usefulness of a labor-force classification based upon a particular week has been much more strongly challenged in connection with the Decennial Census, where the results are available too late to be of current interest and where a classification based upon a longer time period may be in order. In certain undeveloped countries where a large proportion of the population is engaged in its own account work or in family employment, and where the economic fluctuations of more industrialized countries are lacking, the weekly time reference is subject to serious question.

Hour Limits for Determining Employment. The fact that a person working as little as one hour in the week is classified as employed is frequently criticized Many critics seem to feel that a person doing a trivial amount of work in a week ought not to be included in the nation's labor force. This type of difficulty, of course, is encountered at the cutting edge of many classifications, and the provision of tables classifying the employed by hours of work makes it possible to determine the effect of alternative definitions.

A more serious point is that many of the persons who are working short hours are doing so involuntarily and want more work. This problem is discussed later in a consideration of the category intermediate between employment and unemployment.

Fifteen Hour Limit for Unpaid Family Workers. The present lower limit of 15 hours work in the week for inclusion of unpaid family workers among the employed was introduced after considerable experience with qualitative rules to eliminate from the labor force persons with only incidential duties on farms and in family businesses. It is believed to have worked reasonably well in giving greater stability in

this part of the labor force. Of course, the difficulty of securing precise reporting of hours worked, especially for intermittent workers, increases the measurement problems associated with any cutoff of this type.

Part-Time Employment Often a Combination of Working and Looking for Work The need for further analysis of part-time workers has led to a good deal of experimentation since 1946 to determine the reasons for part-time employment Persons working less than 35 hours during the survey week were separated into those who usually worked full time at their current jobs and those who usually worked part time. For those normally on full time, the reasons for their reduced hours during the survey week were ascertained The others were asked whether they preferred and could have accepted full-time employment

These special investigations—which have been conducted quarterly during periods of declining economic activity and less frequently at other times—have been among the most successful of the innovations in the survey.[3] The principal additional categories established thereby have been the number of full-time workers on reduced workweeks because of slack work, material shortages, job turnover, and other economic factors and the number of regular part-time workers who preferred and could have accepted full-time employment. The first of these groups—full-time workers on shortened workweeks—has proved to be a highly sensitive indicator of economic trends, often signalling important changes before they were reflected in either the over-all employment or unemployment estimates The second group—part-time workers desiring full-time work—is somewhat less sensitive to these trends, because it includes a hard core of persons in perpetually unstable occupations such as domestic service, but nevertheless shows significant increases during periods of declining activity. The special studies have also revealed that, regardless of business conditions, the large majority of persons working part time are doing so by choice or for various personal reasons, this is true especially of those reporting only a very few hours of work during the survey week

The diagnostic value of data on part-time employment is somewhat limited because the figures have not been available on a monthly basis and because interpolation between the periodic estimates has often been difficult Serious consideration, however, is being given to obtaining the information monthly. It is noteworthy that the Dominion of Canada has adopted a somewhat modified version of these special

[3] For results of these surveys, see U S Bureau of the Census, *Current Population Reports*, Series P-50, Nos. 7, 12, 17, 18, 21, 25, 26, 28, 33, 34, 46, 52, 53, and 55.

questions for its monthly enumeration and that the additional classifications have become an integral part of the conceptual framework.[4]

DIVISION BETWEEN "WITH A JOB" STATUS AND "UNEMPLOYED" STATUS

Controversy over the classification of certain persons as "with a job" has arisen primarily because of the thought that some persons in this classification might more properly be regarded as unemployed There is comparatively little disagreement about the present classification for those absent because of vacations, illness, or other personal reasons, assuming that they can be separately identified so as not to obscure labor-input analysis Substantial, although not universal, agreement exists also in the case of those idle because of bad weather or labor disputes

Most of the discussion has centered around two of the smaller components—persons on temporary (less than 30 day) layoff and those waiting to report to new jobs scheduled to start within 30 days The rationale for inclusion of these persons with the employed, and the counterarguments for classification as unemployed or even (in the case of those awaiting new jobs) as not in the labor force have been described in great detail elsewhere[5] and will not be repeated here. In recognition of this long-standing dispute, however, the Bureau in 1945 started publishing and displaying in a reasonably prominent position in its releases, separate statistics for these categories, so that they could be treated in accordance with the needs and viewpoints of different consumers. More recently, the amount of detail tabulated for these persons has been expanded considerably, which should expedite regrouping if a change in classification is indicated.

The existence of a substantial miscellaneous group of persons absent from work in any given week has also been a source of some uneasiness. This "catch-all" group includes persons not working for a variety of personal or other reasons, no single one of which is sufficiently important to warrant separate classification. The total has ranged between 300,000 and 500,000 in the past several years and is currently somewhat above the level of a year ago. Present plans are to reexamine the various components of this group and to determine whether any meaningful sub-classifications could be established, if so, they would be shown separately in the statistics.

Although attention has been directed primarily at the categories

[4] Dominion Bureau of Statistics, *The Labour Force.*

[5] See, in particular, Social Science Research Council, *Labor Force Definition and Measurement*, by Louis J. Ducoff and Margaret Jarman Hagood, Bulletin No 56, 1947.

just mentioned, there is some sentiment for a general overhauling of the entire concept of persons "with jobs but not at work." The concept implies a very specific job attachment from which the person was only "temporarily" absent. Aside from the "layoff" and "new" job segments, however, where a 30-day cutoff is specified, the term "temporary absence" has never been defined.

In March 1954, an experimental study was conducted to determine the expected duration of absence for persons in the "with a job" group in the interest of evaluating the concept and possibly tightening the criteria. The results showed, for those reporting, that about 75 per cent expected to be back at work within 30 days of the start of their absence and 82 per cent within 60 days. The proportion whose anticipated absence exceeded 60 days was largest among those not working because of bad weather or for miscellaneous reasons, probably because both these groups include many seasonal workers who ordinarily do not work during the winter months The rules specify that seasonal workers should not be reported as having jobs during the off season, and some need for tightening this provision may be indicated by the results. Absences of substantial duration because of illness were also reported in a number of cases, suggesting some need for clarifying the concept of "temporary" illness.[6]

A further aid in evaluating the concept would be information on the specific nature of the job arrangements of persons in the "with a job" category. Information of this type may be obtained in the future as a byproduct of a system of quality control now being developed by the Bureau. For this purpose, a highly detailed questionnaire may be used by the field supervisory staff in rechecking the survey results.

DIVISION BETWEEN THE UNEMPLOYED AND PERSONS NOT IN THE LABOR FORCE

The greatest amount of controversy over labor force concepts has involved the classification of persons who may frequently cross and recross the line of division between being definitely outside the labor force and being in the market for a job. Some consumers would prefer the inclusion as labor force members of all persons willing and able to work, others would restrict the concept to those actually working or intensively seeking jobs. The concepts adopted in 1940 sought to steer a middle course, establishing some standards of objectivity, without imposing severe restrictions based on intensity of job-seeking efforts and similar criteria and with some recognition of special circumstances

[6] The results are under review by an interagency committee re-examining labor-force concepts.

Although the concepts represented a distinct advance from the standards used previously, experience in their application has demonstrated that there are many borderline situations which continue to be troublesome, if not insoluble. It has been suggested that the notion of a distinct line of demarcation between workers and nonworkers may be fallacious in itself, and that we merely have a continuum reflecting different degrees of association or lack of association with the labor market.

The Bureau has conducted a number of experiments designed to measure the size and composition of the group on the fringe of the labor market which, under current practices, is excluded from the labor force count Although the approach has varied, the usual procedure has been to identify those persons originally reported as nonworkers who had worked or looked for work within the previous month or two, and to establish either the reason for their current inactivity or whether they still desire employment [7] The results have shown that there is a small group—possibly ranging between 300,000 and 500,000 —who resemble the "inactive" unemployed in some respects or who, without any serious departure from current concepts, could reasonably be considered as labor force members This group consists largely of housewives and teen-age boys and girls and appears to increase slightly but not strikingly during periods of business downturn. Unfortunately, there have been no measurements of this type within the past four years; and some current checks, hopefully using a more imaginative approach, are needed

USE OF GROSS CHANGE DATA AS TOOL OF ANALYSIS

The economic significance of changes in employment and unemployment may sometimes be obscured because of the existence of a large group of part-time and intermittent workers who move into and out of the labor force with considerable frequency in the course of a year. Seasonal adjustment of the series alleviates the problem to some extent, but such adjustments are at best only approximations. The availability of data by age, sex, and other characteristics is a substantial aid, although demographic groups, of course, are far from homogeneous.

An important advance in labor force analysis—which has both identified the magnitude of this problem and pointed to a partial solution—is the development of information on "gross changes" in

[7] For summary of these studies, see U. S Bureau of the Census, *Labor Force Memoranda*, Nos. 3 and 4.

employment and unemployment. About 75 per cent of our sample is common from month to month, thus providing a basis not only for determining status at any given time but also for examining changes in status for an identical group of persons These tabulations show that, regardless of how small net changes may be from month to month, millions enter or leave the labor force or shift between an employed and an unemployed status [8]

By these means it is possible to isolate, to some extent, changes of prime economic significance from those of perhaps secondary importance When unemployment is rising, for example, one may determine the extent to which the increase is due to job losses and the amount attributable to the entry of housewives, students, and others into the labor market in search of jobs Conversely, when unemployment is dropping, job accessions and withdrawals from the labor market can be separately identified Even where there is little apparent over-all change, gross movements may reveal certain significant underlying developments

These data have been highly useful in interpreting month-to-month trends as well as in illustrating the dynamic nature of labor market activity. It is hoped that we may be able to expand the tabulations to cover changes in hours worked and other special subjects in addition to employment status.

Although gross change tabulations to date have been confined to month-to-month comparisons, it will shortly become possible to extend this type of analysis to year-to-year changes. Time and space do not permit coverage of all the analytical possibilities that could be exploited by these means Since year-to-year differences usually represent more permanent changes in status, it is clear that such matters as the pattern of entry into, and retirement from, the labor force, basic occupational and industrial shifts, and many related subjects could be thoroughly explored for perhaps the first time Both on an annual and a monthly basis, there also remains the relatively untapped area of the motivation for these changes in status, which presumably could be pursued through supplementary questions. Here again, however, much experimentation would be needed in order to obtain meaningful results

RECONCILIATION OF CENSUS SERIES WITH OTHER GOVERNMENT SERIES

In addition to the Census figures on the total numbers employed and unemployed, there are three other current government series in

[8] The magnitude of these gross movements may be exaggerated somewhat because of some inconsistency in reporting for identical persons from one month to the next where no change in status has

the same general field; the Bureau of Labor Statistics (BLS) data on employees of non-agricultural establishments; the Agriculture Marketing Service (AMS) data on farm employment, and the Bureau of Employment Security (BES) data on insured unemployment It is generally agreed that the series serve different purposes and thus supplement one another in a very important fashion Perhaps the best indication of their joint usefulness is the fact that three of the four series are being issued currently in a combined release by the Departments of Labor and Commerce.

The differences between the Census series and the other three can be explained in considerable part by technical considerations connected with the fact that the Census series is based upon information collected directly from households, whereas information from the other three series is provided by employers The major conceptual difference between the Census and the other employment series arises from the treatment of workers with more than one job in the reporting period Both the BLS and AMS statistics are collected from employers and essentially represent a count of different jobs, whereas the Census counts a person only once—in his major job—regardless of the number of jobs held Another major difference from the BLS series lies in the treatment of persons with a job but not at work, BLS counts these persons only if they received pay for the reporting period Comparisons between Census and AMS are also affected considerably by the inclusion of children under 14 years of age in the latter series.

Tests which have been applied to the Census series to measure the size of the conceptual differences between Census and BLS data have given some information on the number of dual job holders and on the numbers of persons on unpaid absences who would be excluded from the BLS series if payroll data from companies do exclude all persons receiving no pay in a particular payroll period [9] The conclusion emerging from these studies is that conceptual differences are definitely not the only and perhaps not even the major problem in this field and that future efforts at reconciliation should give at least equal attention to sampling and measurement techniques It is hoped that after necessary studies of sampling and other errors in both series have been completed, there may be a better understanding of the relationships between them, and their joint usefulness may be increased.

actually occurred Generally, the net of gross movements between two categories will be more accurate than the gross movements themselves, although even the latter, in the case of unemployment for example are reasonably consistent with data from other sources.

[9] Some consideration is being given to obtaining pay status on a regular or at least frequent basis for persons with jobs but not at work during the survey week

In the case of unemployment, the main emphasis has been upon the development of adjustment factors which would compensate for differences in coverage between the Census estimates and the BES figures on insured unemployment. This procedure has achieved fairly satisfactory results.[10] The problem of reconciliation here is most troublesome at a time of large-scale withdrawals from the labor force, when large numbers of women may be ready to continue the jobs which they held during a period of emergency but are not seeking any other work.

SUMMARY

In summary, the past 12 years have involved a great amount of research under committeee auspices to sharpen labor force concepts, to develop better procedures for collecting the data, and to provide many types of supplementary information. Advances in methods, involving the most modern techniques of sampling and of electronic tabulation, have also been noteworthy That much remains to be done is very clearly shown in the report of the Special Advisory Committee on Employment Statistics appointed by the Secretary of Commerce,[11] and in the coming months we feel sure that progress will be made along the following lines:

1. Intensive study of concepts with the guidance of both Government and private advisory groups.
2. Instensive research on questionnaires, forms, etc., to see what advances in objectivity and reductions in non-sampling errors can be made.
3. Greater emphasis on quality control procedures to insure maintenance of proper standards of performance.

---

[10] See *Comparison of Census Bureau Estimates of Unemployment and Insured Unemployment Statistics*, a statement submitted by the Bureau of the Census and the Bureau of Employment Security to the Joint Congressional Committee on the Economic Report, February 1954.

[11] "The Measurement of Employment and Unemployment by the Bureau of the Census in its Current Population Survey," Report of the Special Advisory Committee on Employment Statistics, August, 1954.

EXAMINATION OF TWO SOURCES OF ERROR IN THE ESTIMATION OF NET INTERNAL MIGRATION*

DANIEL O. PRICE
University of North Carolina

Two sources of errors in the estimation of net internal migration are examined. First, errors arising from the use of a single set of survival rates in all 48 states are examined by estimating net migration in ten states using both a national life table and state life tables. It is estimated that the median error arising from this source is about 14 per cent of the estimate of net migration.

The second source of error examined is underenumeration of the population The "built in" correction factor of Census survival rates is demonstrated algebraically, and one approach is made to estimating the magnitude of errors from this source when the assumptions are not justified. It is estimated that about one-third of the estimates of net migration are in error by 25 per cent or more due to the effects of underenumeration.

These two estimates of error are quite rough and the main conclusion to be derived is that small relative differences in estimates of net migration should be interpreted with extreme caution.

A GREAT many man-hours have been spent in measuring—or rather in estimating—net migration for various areas and population groups The methods used are generally familiar, that is, the natural increase method and the survival rate or residual methods, survival rates being computed either from life tables or Census figures and either the forward, reverse, or combined procedure being used [2]. In both the natural increase method and the survival rate method an expected population is computed and compared with an observed population and the difference is attributed to migration. This difference which is attributed to migration, however, is really the result of several factors. In nearly all cases migration is the largest single factor present so it is proper to think of the other factors as causing errors in the estimate of migration. We would like to know the relative importance of these various sources of errors and would like to be able to say that for a certain method of estimating net migration the standard deviation of the errors is a certain amount, say, 10 per cent. That is, we could say that about two-thirds of our estimates were within plus or minus 10

* Revision of a paper presented at the Annual Meeting of the Population Association of America, Charlottesville, Virginia, May 8, 1954.

per cent of the true value. Given such an estimate of the variance of the errors we would like to be able then to break up this variance into its component parts somewhat as follows:

$$\sigma^2 = u^2 + s^2 + f^2 + \cdot \quad \cdot + \epsilon^2 \tag{1}$$

Where

σ^2 is the total error variance

u^2 is the amount of the total error variance due to errors of under-enumeration

s^2 is the amount of the total error variance due to errors in survival rates

f^2 is the amount of the total error variance due to using the forward method rather than some other, etc , and

ϵ^2 is a residual amount of unexplained error

If it were possible to assign values to each term in equation (1) we would then have an objective basis for planning further research to improve the accuracy of estimates of net migration. That is, it would seem logical to attack that source of error which contributed most to the total error variance. For example, if 90 per cent of the total error variance came from errors in survival rates, then this would be the logical point on which to work for improvement rather than on some other source of error that accounted for only five per cent of the total error variance. This ideal, however, isn't attainable

In the first place, the model given in equation (1) is oversimplified because it assumes that the sources of error are independent of each other. This is a simplifying assumption and may not be reasonable. It is imaginable that those areas where underenumeration is worst might be the same areas for which the survival rates are most in error. If these sources of error are correlated, then we must add a cross-product term involving the correlation coefficient between these two sources of error If two sources of error are negatively correlated, they tend to compensate for each other and the total error is reduced; but if they are positively correlated, the total error is increased.

In addition to the fact that equation (1) is oversimplified, we are not able to evaluate the terms of it, much less determine the amount and direction of correlation between the sources of error. However, it is felt that efforts to evaluate the terms of (1) or a modified version of it are worth the attention of demographers The balance of this paper is a report of efforts to evaulate or show the importance of two sources of error in estimating net migration by the survival rate method.

Let us first look at the effect of ignoring regional differentials in mortality. Whenever we use the same set of survival rates for all states we are ignoring differences in rates of mortality from one state to another. For example, we assume that the average rate of mortality for the 20–24 age group holds for all states When we apply this average rate of mortality in a state that actually has a higher rate of mortality, we expect more survivors at the end of the decade than there really are, and we attribute the deficit to out-migration The reverse happens in a state with the mortality rate lower than average.

In order to evaluate the magnitude of errors arising from this source, we have computed estimates of net migration for ten states by two different methods The ten sample states were selected by taking every fifth state from a regional listing of the 48 states. For each of these ten states net migration of native white males was estimated using survival rates taken from a United States life table for white males [3], the same set of survival rates being used for all ten states. These estimates of net migration are shown in column 1 of Table 1 Estimates of net migration were then made using survival rates taken from individual state life tables [3] These estimates are shown in column 2 of Table 1 and take into consideration state differences in rates of mortality

TABLE 1

ESTIMATES OF ERRORS DUE TO USING A SINGLE SET OF SURVIVAL RATES IN THE ESTIMATION OF NET INTERNAL MIGRATION OF NATIVE WHITE MALES IN TEN SAMPLE STATES, 1930–1940

State	Estimated Net Migration Using Survival Rates From			Estimated Error (2) − (1)	Error as Per Cent of Net Migration (4) ÷ (3)
	U S life table, 1930 (1)	State life tables, 1930 (2)	U S Census (3)	(4)	(5)
Ariz	7,790	15,553	6,375	7,763	121 8
Ark.	−39,425	−43,317	− 44,302	3,892	8 8
Fla	106,859	108,721	102,672	1,862	1 8
Ill.	−15,159	− 9,522	− 38,097	5,637	14 8
Mo	− 6,383	−12,167	− 20,405	5,784	28 3
Mont.	− 3,566	− 3,360	− 4,878	206	4.2
Nev	8,200	9,235	7,854	1,035	13 2
N C.	− 5,561	− 6,426	− 13,541	865	6 4
Penn.	−94,314	−71,472	−122,307	22,842	18 7
R. I.	2,574	3,225	911	651	71 5

In getting these two estimates of net migration between 1930 and 1940, life tables for 1930 were used. It would have been desirable to use an average of the 1930 and 1940 life tables However, our interest is in the difference between the two methods, and it is assumed that the difference between the two methods would be relatively constant regardless of whether 1930 life tables were used, 1940 life tables were used, or an average of the two. There is some reason to believe that state differences in mortality are decreasing, and if this is true, errors arising from this source will decrease.

It is not felt that life table survival rates give as good estimates of net migration as Census survival rates [1], therefore column 3 of Table 1 gives estimates of net migration for the ten sample states computed using Census survival rates. We assume that the differences between corresponding figures in columns 1 and 2 are good estimates of the errors arising from the use of a single set of survival rates, and these differences are shown in column 4. We can express the absolute values of these errors as percentages of the number of migrants, and these percentages are shown in column 5.

This method of expressing the error (as a percentage of the net migration) can, theoretically, lead to division by zero in the case where there is no net migration However, none of the 48 states actually has zero net migration and the use of positional summarizing measures will tend to offset the effect of any extreme values arising from a small net migration. One might expect that the largest percentage errors would tend to occur in those states having the smallest amount of net migration, but a rank order correlation gives little indication of such an association. (Kendall's *Tau* for the sample is .24 The probability of a sample *Tau* being equal to or larger than this when there is no association in the universe is about .32)

Utilizing the percentage errors of column 5 in Table 1 it is possible to make an estimate of the median percentage error and to set confidence limits on this estimate, the confidence limits being set without any assumptions as to the form of the distribution [4]. Our estimate of the median percentage error is 14 0 per cent and the 89 per cent confidence limits are 6 4 per cent and 28 3 per cent. Some critics might object to using Census survival rate estimates of net migration as a base when the errors are based on life table survival rates. The estimates of net migration based on state life tables can be used as the base in computing the percentage error. Such computations lead to an estimated median error of 16 8 per cent with 89 per cent confidence limits of 9.0 per cent and 47.5 per cent.

These findings lead us to the conclusion that half of our estimates of net migration of white males are in error by more than 14 per cent. At least it would seem safe to say that about half of our estimates of net migration include an error greater than a figure of the order of 10 to 20 per cent.

We must keep in mind that these figures are only for white males during the 1930–1940 decade and are rough estimates at best. Whether similar results would be gotten with other population groups and in other decades is unknown. The important conclusion to be derived is that the errors resulting from using a single set of survival rates for all states are not negligible. This finding also implies that most present estimates of net migration which have been computed using a single set of survival rates for all states are by no means exact and due caution should be exercised in comparisons and interpretations.

We must also remember that all of our findings here have been on the basis of total net migration rather than net migration by age groups. The errors in the estimation of net migration by age groups are possibly larger than in the estimation of total net migration, this fact should be kept in mind in using figures on net migration by age.

Another important source of error in estimates of net migration is underenumeration of population. As was mentioned above, we usually estimate net migration by computing an expected population, comparing it with an observed population, and assuming that the difference is due to migration. A good deal of attention has been given to methods of computing the expected population, but little attention has been given to the effects of errors in enumeration of the observed population The second part of this paper makes an effort to examine the relative importance of underenumeration as a source of errors in estimation of net migration. The term underenumeration is used here to include misstatements of age and other misclassifications since from a practical point of view it is impossible to separate them. For example, if we get an undercount in a particular age group, there is no way to know how much of this undercount is due to people who were not counted at all and how much of it is due to people who should have been counted in this group but were counted in some other age group Because of this misclassification by age and other categories it is possible to get an overcount in an age group, but we will use the general term *underenumeration* to include all of these situations.

Information on underenumeration is not available in sufficient detail for us to estimate empirically the magnitude of the errors from this source as was done above for survival rates. It is possible, however,

to demonstrate algebraically the part that underenumeration plays in estimates of net migration. If we let

A = U. S. native population in age group x at end of decade
B = U. S native population in age group $x-10$ at beginning of decade
C = state native population in age group $x-10$ at beginning of decade
D = state native population in age group x at end of decade

$$A/B = \text{survival rate, } R$$

Net migration in state i is

$$M_i = D_i - \frac{A}{B} C_i$$

A, B, C, and D are measured with errors a, b, c, and d so that we know only

$$(A - a), \qquad (B - b), \text{ etc., or}$$

$$A\left(1 - \frac{a}{A}\right), \quad B\left(1 - \frac{b}{B}\right), \quad C\left(1 - \frac{c}{C}\right), \quad D\left(1 - \frac{d}{D}\right)$$

Our estimate of net migration in state i is

$$M_i' = D_i\left(1 - \frac{d_i}{D_i}\right) - \frac{A}{B}\,\frac{\left(1 - \frac{a}{A}\right)}{\left(1 - \frac{b}{B}\right)}\,C_i\left(1 - \frac{c_i}{C_i}\right) \qquad (3)$$

We can look at this equation by parts in order to see the effect of various sorts of underenumeration Let us first look at the middle of the second term, that is

$$\frac{\left(1 - \frac{a}{A}\right)}{\left(1 - \frac{b}{B}\right)}$$

This represents the amount of error in the survival rate If the underenumeration in the United States population in the age group at the beginning of the decade is the same as the underenumeration in the same cohort at the end of the decade, then there is no error in the sur-

vival rate from this source. However, the rationale in using Census survival rates for estimating net migration is not that $(a/A)=(b/B)$ but that the undercount or overcount of an age group in a particular state will be approximately the same as the undercount or overcount of the corresponding group in the U. S. Put into symbols this is

$$\left(\frac{a}{A}\right) = \left(\frac{d}{D}\right) = e$$

$$\left(\frac{b}{B}\right) = \left(\frac{c}{C}\right)$$

If these conditions hold, then formula (3) becomes

$$M_i' = \left(D - \frac{A}{B} C_i\right)(1 - e)$$

This demonstrates the "built in correction" of Census survival rates. However, even when our assumptions hold, the estimate of net migration is in error by the same percentage as the underenumeration of the cohort at the end of the decade. This is an error in the estimation of the absolute number of net migrants Horace Hamilton has pointed out to the writer, however, that when the number of net migrants is related to the population at the end of the decade in order to get a "migration rate," the error cancels out and the *rate* is correct. Dr. Hamilton has also demonstrated that if a reverse solution, or "revival method," is used, the error still cancels out when a rate is computed.

The real problem arises in an attempt to estimate the extent to which the underenumeration of an age group in a state departs from the underenumeration in the corresponding age group in the entire United States. The logical approach to this problem is through the study of the variation among states in underenumeration for particular age groups. We do not, however, have complete information on underenumeration by age and by states (If we had this, of course, we could adjust our estimates of net migration and avoid this source of error) The first approach to this problem was to assume that the coefficient of variation of underenumeration would be relatively constant for most age groups. This is to say that the greater the underenumeration of an age group for the total United States, the greater will be the variations of states about this average underenumeration, and that the ratio of the standard deviation of the variations to the average underenumeration will be approximately constant from one age group to another. It is possible to make some tests of this hypothesis utilizing scattered data on under-

enumeration. The results of such tests using the available data are shown in Table 2.

An examination of this Table does not support the assumption outlined above. Three out of the four groups, however, have approximately the same standard deviation regardless of the average level of underenumeration. The fourth group, Negro males of draft age, has a much larger standard deviation but this might be a function of the fact that

TABLE 2

MEANS AND STANDARD DEVIATIONS OF PER CENT UNDERENUMERATION BY STATES FOR FOUR POPULATION GROUPS

Population Group	Mean Per Cent Underenumeration	Standard Deviation	No. of States Covered
Whites under 5 years of age, 1940*	6.10	2 37	47
Nonwhite under 5 years of age, 1940*	15.42	2.96	16
Males 21–35 years of age inclusive, 1940†	2.00	2 83	48
Negro males 21–35 years of age inclusive, 1940†	15.87	7.08	23

* *Sixteenth Census of the United States, 1940 Population* Differential Fertility, 1940 and 1910. Table A-2

† Daniel O Price, "A check on underenumeration in the 1940 census," *American Sociological Review*, XII (February 1947), 44–9

the 23 states used in computing this figure were those with the largest Negro populations. Since this is not necessarily the case, however, the figures of Table 2 do not give unqualified support to the hypothesis that the standard deviation of underenumeration is a constant.

Let us assume, however, that the three smaller standard deviations of Table 2 give us some basis for estimating the standard deviation of errors of underenumeration among states. Using this assumption, we can estimate the effects of errors of underenumeration. We begin by making the following substitutions in equation (3).

$$\alpha_i = \left(1 - \frac{d_i}{D_i}\right)$$

$$\beta_i = \left(1 - \frac{c_i}{C_i}\right)$$

$$K = \frac{1 - \dfrac{a}{A}}{1 - \dfrac{b}{B}}$$

$$\frac{A}{B} = R$$

Then our estimate of net migration is

$$M_i' = D_i\alpha_i - RKC_i\beta_i$$

Where

$$M_i = D_i - RC_i$$

The error due to underenumeration is

$$M_i - M_i'$$

We express this error as a proportion of the population at the end of the decade thus.

$$E_i = \frac{M_i - M_i'}{D_i}$$

We assume that this error is uncorrelated with the amount of migration and therefore wish to determine the standard deviation of E and relate it to the mean proportion of the population classed as migrants and get a coefficient of variation.

$$E_i = \frac{(D_i - RC_i) - (D_i\alpha_i - RKC_i\beta_i)}{D_i}$$

$$= (1 - \alpha_i) - TK\frac{C_i}{D_i}\left(\frac{1}{K} - \beta_i\right) \qquad (4)$$

We could substitute typical values of α, β, R, K, C, and D in this formula and make an estimate of the average value of E; but it seems more useful instead to estimate the variation in E, and this we will proceed to do.

The ratio C_i/D_i is the ratio of the population aged $x-10$ in a state at the beginning of a decade to the number left in a state in the same cohort at the end of the decade. Without migration this term will always be less than 1.0, but with migration it will vary over a range from about .8 to 1.1 for those age groups most affected by migration.

Let us call this ratio P_i and assume that it has a standard deviation of .06 since this is consistent with the range assumed above.

The term $(1-\alpha_i)$ is simply d_i/D_i or the underenumeration of D_i expressed as a proportion of the population. Let us call this term δ Table 2 gives us reason to assume that the standard deviation of this term is about 027. Let us assume a normal distribution.

The term $[(1/K)-\beta_i]$ has the same variation as β_i which has the same variation as c_i/C_i Let us call this term γ_i. We are assuming that the variation of c_i/C_i is the same as d_i/D_i and that c_i/C_i is probably correlated with d_i/D_i We are also assuming that both of these terms are independent of C_i/D_i We can now write that

$$E_i = \delta_i - RKP_i\gamma_i$$

The variance of E, $\sigma^2(E)$, can be written

$$\sigma^2(E) = \sigma^2(\delta_i) + RK\sigma^2(P_i\gamma_i) - 2RKr\sigma(\delta_i)\sigma(P_i\gamma_i)$$

where r is the correlation between δ_i and $(P_i\gamma_i)$. Since P_i and γ_i are independent, this can be written as

$$\sigma^2(E) = \sigma^2(\delta_i) + RK\sigma^2(P_i)\sigma^2(\gamma) - 2RKr\sigma(\delta_i)\sigma(P_i)\sigma(\gamma_i) \qquad (5)$$

We now have the variance of the errors due to underenumeration expressed in terms of the variance of the individual components and all that remains is to evaluate the expression. Substituting the assumed values in expression (5) gives us

$$\begin{aligned}\sigma^2(E) &= .000729 + RK(.0036)(\ 000729) - 2RKr(.027)(.06)(.027)\\ &= .000729 + RK(\ 000006262 - 2r\ .00004374) \qquad (6)\end{aligned}$$

We can assume various values for R, K, and r and evaluate equation (6). For most age groups the true survival rate, R, will be between .85 and 95 It also seems reasonable to assume that the ratio, K, of the computed survival rate to the true survival rate would be between 9 and 1 1 The correlation between δ_i and $P_i\gamma_i$ is probably between 0 and 1. If $\sigma(E)$ is computed for all combinations of R, K, and r within the ranges specified above, it varies only from .025 to .027 Thus we have an estimate of the standard deviation of the errors arising from underenumeration.

From this we can estimate that about one-third of the errors in estimates of net migration arising from underenumeration are more than 2½ per cent of the population in the state at the end of the decade Actually the errors are larger than this because this is only the variation of the errors about some average error. However, if we make the as-

sumption that the average error is zero we still have errors of the magnitude described above.

If we relate this estimated standard deviation of errors to the mean net migration, we can get a coefficient of variation. However, we must remember that these estimates of underenumeration are based on age groups and therefore we must relate this estimated standard deviation to mean net migration by age Since one of our estimates of underenumeration comes from the 21–35 age group, let us relate our estimate of error to the mean net migration of the 20–35 age group. This age group has one of the highest rates of net migration; therefore, we will tend to get an estimate of minimum error. During the 1930–40 decade the net migration in the 20–35 age group was between 9 and 10 per cent of the 1940 population for those states having net out-migration as well as for those states having net in-migration in this age group. If we use 10 per cent as our estimate of the mean net migration in this age group, the coefficient of variation is

$$\text{C. V.} = \frac{2.5}{10} = 25 \text{ per cent}$$

This suggests that about one-third of our estimates of net migration (by the survival rate method) for specific age groups are in error by more than 25 per cent due solely to the effects of underenumeration.

If our results from the examination of these two sources of error are approximately correct, we see that the errors arising from underenumeration are at least as important as errors arising from the use of a single set of survival rates for all states. Thus in addition to improving estimates of survival rates, we should attempt to devise ways to estimate underenumeration and correct for the errors introduced. The lack of data and the difficulty of obtaining the necessary data complicates the task; but if estimates of net migration are useful, surely ways can be devised to make adjustments for errors arising from these sources.

In conclusion we might look briefly at the two sources of error together—that is, errors arising from the use of a single set of survival rates for all states and erorrs arising from underenumeration Are errors from these two sources correlated? If they are negatively correlated, then they tend to act in such a way as to reduce the total errors of estimate of net migration. On the other hand, if they are not correlated, the variance of the errors from the two sources will be additive If they should be positively correlated, the total error variance will be greater than the sum of the error variances from the two sources There seems

to be little reason why these two sources of error should be correlated at all, although it might be possible to make a logical case for a positive correlation between them. At any rate we can be fairly certain that the methods most frequently used to estimate net migration give us estimates that are likely to include an appreciable percentage error. For this reason caution must be exercised in interpreting such figures and small relative differences between states and decades should not be taken too seriously.

REFERENCES

[1] Hamilton, C. Horace, and Henderson, F. M., "Use of the survival rate method in measuring net migration," *Journal of the American Statistical Association*, 39 (June 1944), 197–206.

[2] Siegel, Jacob S., and Hamilton, C. Horace, "Some considerations in the use of the residual method of estimating net migration," *Journal of the American Statistical Association*, 47 (September 1952), 475–500.

[3] U. S. National Resources Committee, *Population Statistics*, Part 2 State Data Washington U S Government Printing Office, 1937

[4] Walker, Helen M., and Lev, Joseph, *Statistical Inference* New York: Henry Holt and Co., Inc., 1953, 440.

THE REDESIGN OF THE CENSUS CURRENT POPULATION SURVEY

MORRIS H. HANSEN, WILLIAM N. HURWITZ,
HAROLD NISSELSON, AND JOSEPH STEINBERG*
Bureau of the Census

In February 1954 a redesign of the Current Population Survey was introduced that provided for a more efficient system of field organization and supervision as well as some advances in methods. The sample is now spread over 230 areas instead of 68 areas with the same number of households as heretofore. A composite estimation procedure has been introduced which reduces the sampling variability for most estimates Also, there has been a considerable reduction in the variance of variance estimates made from the sample. A statistical quality control program has been introduced to help insure results of consistently acceptable quality. Problems arose in the process of shifting from one design to the other that resulted in some significant differences between the new and old samples for a few of the estimates, especially unemployment. Apparently response errors were the principal source of difficulty, and it was possible to take steps to bring the results within sampling error range. Work on the measurement and control of response errors is being expanded.

INTRODUCTION

IN FEBRUARY 1954, the Bureau of the Census introduced a redesign of Current Population Survey, from which information on employment, unemployment, and related data are compiled each month. Information on other topics, such as income distribution, family characteristics, marital status, migration, and education, are compiled less frequently. Since 1943, the estimates had been made from a sample of households in 68 primary sampling units spread throughout the United States. The main features of this sample design, which has been in operation for more than a decade, have been described elsewhere,[1] they will not be reviewed here, except where it is necessary to point up some of the principal changes introduced. It is to be noted, however, that the sample was set up as a general purpose sample, and the sample of areas is used for a monthly retail trade survey and numerous special surveys as well as for the monthly labor force measurements.

---

* The work reported was the joint work of sampling and other staff members in the Bureau of the Census. This paper was presented at the Annual Meeting of the American Statistical Association at Montreal in September 1954

[1] Hansen, Morris H, Hurwitz, William N., and Madow, William G, *Sample Survey Methods and Theory*, Vol. I, Chapter 12 B (prepared by J. Steinberg), New York John Wiley and Sons, 1953, 559–82.

In summary, the redesign provides for a more efficient system of field organization and supervision, takes advantage of a more widespread sample that is made feasible when the requirement of a full-time supervisor in each sample area (which was accepted as a condition in the original design) is removed, and introduces some other advances in methods. As a result it is possible to provide more information per unit of cost, to increase the accuracy of published statistics, and perhaps to make regional estimates for summary characteristics. Also, variance estimates made from the sample are substantially improved.

The main features of the new design that differ from the old are that the sample is now spread over 230 instead of 68 primary sampling units but with approximately the same total number of households included in the sample,[2] the supervisory organization has been changed, the philosophy and methods of control of the field work have been modified, and a new estimation procedure that takes advantage of a rotating sample in estimating time series has been introduced. Also, the tabulations and estimates are now prepared on a high-speed electronic computer, the Univac

The fundamental principles of optimum design of a survey with multistage probability sampling are the same in the two surveys, but the redesigned survey incorporates some advancements in the application of the principles The over-all guiding considerations in the design of a sample survey are to select the sample and carry through the collection, tabulating, and estimating operations in such a way as to provide the desired information of the required accuracy (taking account of both sampling and nonsampling errors) within the necessary time limits and at near the minimum possible cost. We have also considered it important to be able to estimate the sampling errors of the estimates from the sample itself, and also, to the extent feasible, to estimate the nonsampling errors.

The principal new features of the survey design and their implications are discussed more fully in the material that follows

---

[2] A total of 24,000 to 26,000 dwelling units or other living quarters are designated for inclusion in the sample each month Completed interviews are obtained from about 20,000 to 22,000 households. Of the remainder, about 500 to 1,000 are households from which information should be collected but is not because the respondents are not found at home after repeated calls, or are temporarily absent, or are unable to be interviewed for other reasons The other 2,500 to 3,500 designated units represent those found to be vacant, or occupied by persons with residence elsewhere, or otherwise not to be enumerated The over-all sample size varies over time partly because of chance but also because of the growth of the population and the creation of new households Every 2 or 3 years as the sample expands with population growth, it is necessary to decrease the sampling ratios slightly in order to keep the workload at the average of roughly 25,000 designated units prescribed under the survey budget

PRINCIPAL INNOVATIONS IN THE NEW SURVEY DESIGN

Supervisory organization and its relationship to optimum design. In the original design of the 68-area sample a condition was imposed that a district office should be located in every primary sampling unit included in the sample An office was staffed with a supervisor, at least one full-time clerical assistant, and, in a few of the offices, with additional supervisory and clerical personnel.

The requirement of a full-time supervisor in each p s.u. was imposed in an effort to insure that the work was carried out substantially in accordance with the specifications, not only in order to have a valid probability sample, but particularly in order to control response or measurement errors. It was felt that the proximity of the supervisors to the part-time interviewers would permit almost daily contact with the interviewers and therefore would provide effective means for supervisory control as well as a means for giving formal training to interviewers each month With this restriction and with the available financial resources, a sample of about 21,000 interviewed households spread over 68 areas approximated the optimum design

If the supervisory restriction had not been imposed, the optimum design would have involved spreading the sample over a much larger number of primary sampling units, provided only that performance could be controlled satisfactorily so that response and other nonsampling errors would not be increased

More recently, evaluation of the one-supervisor-per-p s u. type of organization led us to the conclusion that we were paying a considerable overhead cost for each field office, and that we were not making as effective use as we might of the supervisory staff Frequently, supervisors were doing interviewing and clerical operations because the supervisory aspects of the work did not constitute a full-time job. We concluded that the ratio of supervisory offices to interviewers was excessive, and that by introducing appropriate statistical quality control principles in the control of the field work, a more effective organization could be established Consequently, we decided to reduce the number of supervisory offices by about 50 per cent, with a smaller reduction in rent, clerical staff, and other overhead costs associated with the offices The supervisory personnel would be reduced only by about 25 per cent. This made it possible to retain a smaller number of higher-grade persons and from the savings in overhead, to pay for the additional travel involved in spreading the sample into a much larger number of primary sampling units.

As before, this field organization is also used for the monthly retail trade survey, which is taken in the same areas, and for other field work.

Sample selection. The expansion of the sample from 68 to 230 areas reduced markedly the sampling error of most of the important estimates of aggregates, and may make it possible to publish summary statistics for several geographic regions The expansion of the number of p s u.'s reduced considerably the component of the variance due to the sample of p.s.u.'s, and made little or no change in the component of the variance due to the sample within p s.u.'s. The between-p.s.u. variance term dominated the variance for most summary items in the original sample, and thus the gains through the reduction of this contribution are worthwhile. Table 1 shows approximate estimates of the between-p s.u. and the within-p s u. contributions to the variance for the new sample for January 1954 For many items, the estimated between-p s.u contribution to the variance for the 68-area design was more than five times as large as that for the 230-area design; the within-p.s.u. contribution for the 68-area design was about the same as for the 230-area design

A particularly important gain from the increase in the number of p.s.u.'s is the increase in the precision of variance estimates. With the 68-area sample, variance estimates made from the sample itself were subject to relatively wide sampling fluctuations because of a number of factors involved in stratified sampling from non-normal populations,[3] and especially because the estimates were based on a small number of degrees of freedom. With the 230-area sample, the number of degrees of freedom has been increased sufficiently to yield variance estimates of much greater precision.

In the new design, as in the old, each p s.u. was defined to be as heterogeneous as feasible subject to a restriction on its total area. Thus each p.s.u. included one or a few adjoining counties. In the new sample, however, each standard metropolitan area either constitutes a self-representing area or a separate p s.u in order to facilitate making estimates for the aggregate of all metropolitan areas from the present sample, and to facilitate the expansion of the sample when data for separate metropolitan areas are desired.[4]

More recent data were used in defining the strata into which the primary units were grouped although only the total population counts

---

[3] For a discussion of this problem, see Hansen, Morris H , Hurwitz, William N , and Madow, William G.: *Sample Survey Methods and Theory*, New York John Wiley and Sons, 1953, Vol I, Chapter 10, 431-35.

[4] Standard metropolitan areas were established subsequent to the original sample design. In the original design the self-representing p.s.u's approximated metropolitan districts, as then defined.

TABLE 1

ESTIMATED REL-VARIANCES (COEFFICIENTS OF VARIATION SQUARED) OF SELECTED LABOR FORCE ITEMS FOR 230-AREA CPS SAMPLE, JANUARY 1954

Item	Total Rel-Variance (v^2)	Between-P.S.U. Contribution (b^2)	Within-P.S.U. Contribution (w^2)
Civilian labor force divided by population 14 years old and over	.000031	.000010	.000021
Agricultural employment divided by farm population 14 years old and over	0019	00026	0016
Nonagricultural employment divided by population 14 years old and over	000057	.000019	.000038
Unemployed divided by population 14 years old and over	0015	.00036	.0012

from the 1950 Census were available in time for use in stratifying the primary units. The primary strata of the 230-area design were made more nearly equal in size in terms of aggregate populations than in the 68-area design. One primary unit was included in the sample from a stratum, as before, and the primary units included in the sample were selected with probabilities proportionate to population (the 1950 population in the case of the new sample) In the selection of the sample of primary units restrictions between strata were used to increase the geographic stratification of the sample,[5] that is, to decrease the probability of obtaining geographically contiguous primary units.

It is to be emphasized that drawing the sample on the basis of more recent information for stratification and sample selection may result in some reductions in sampling error, but redrawing the sample with up-to-date information is not essential in order to include units in the sample with known probabilities An area probability sample drawn in 1943 provides a probability sample of areas, and consequently of people, retail stores, or other units that can be associated with areas both in 1943 and in 1954, moreover, such a sample would continue to provide a probability sample in future decades provided that the areas can be uniquely identified over such a period and if the existence of a sample in an area does not change the character of the area

[5] Goodman, Roe, and Kish, Leslie, "Controlled selection—a technique in probability sampling," *Journal of the American Statistical Association*, 45 (1950), 350-72.

Up-to-date information from censuses or other sources can be used in the selection of a new sample to provide a more efficient probability sample, but the principal gains from such information can also be made with an area sample drawn at an earlier time by appropriate introduction of such new information in the estimation procedure. This, in fact, was done in making the estimates from the 68-area sample after the 1950 Census data became available The use of more recent information in selecting a new sample may result in some further gains. However, a complete change in the sample would not have been worth the high cost of making the change unless the other gains to be achieved made the changeover worthwhile, as was the case in this particular instance

We emphasize these points because there has been some misunderstanding concerning them Thus, it has sometimes been stated that use of 1940 Census data for stratification and for determining probabilities of selection invalidated the 68-area sample for subsequent periods. In fact, the principal factor that reduces the sampling error in the new 230-area design as compared with the 68-area sample is the increase in the number of p s u.'s. For example, even though the 68-area sample was selected on the basis of 1940 Census data, and the 230-area sample used more recent information, we can estimate 1940 Census characteristics with a smaller variance from the 230-area sample. Table 2 compares some specific estimates for 1940 made from the 68-area sample (selected in 1943) and the 230-area sample (selected in 1952).

The subsampling in the 230 areas was essentially the same as in the 68-area sample, except that 1950 Census data were used in defining the sample segments.

Rotation of the sample One feature of the subsampling which is present in both the old and new designs has an important bearing on the estimation procedure introduced in the new sample This feature involves changing a part of the sample each month to avoid a decline in respondent cooperation (which may happen when a constant panel is interviewed indefinitely) and to reduce the variances of sample estimates under certain circumstances.

To accomplish this rotation, eight systematic subsamples (rotation groups) of segments are identified for each sample A given rotation group is interviewed for a total of eight months, divided into two equal periods. It is in the sample for four consecutive months one year, leaves the sample during the following eight months, and then returns for the same four calendar months of the next year. It is then dropped from the sample. In any one month one-eighth of the sample segments are

TABLE 2

LABOR FORCE STATISTICS FROM COMPLETE CENSUS COMPARED WITH ESTIMATES* FROM 230-AREA AND 68-AREA SAMPLES BASED ON COMPLETE CENSUS TOTALS FOR THE SAMPLE COUNTIES. APRIL 1940

(in thousands)

Employment status	Complete census	230-area sample	68-area sample
Total labor force	52,790	52,850	52,600
Employed.	45,170	45,180	45,060
Farmers and farm managers	5,140	5,190	5,100
Unemployed .	7,620	7,670	7,540
Not in labor force	48,310	48,250	48,500

* Involve approximations to the race-residences and age-sex ratio estimates actually used in CPS, based on 1940 Census data The complete census data are not available in sufficient detail to make identical ratio estimates possible

in their first month of enumeration, another eighth are in their second month, etc., with the last eighth in for the eighth time (the fourth month of the second period of enumeration). Under this system, 75 per cent of the sample segments are common from month to month and 50 per cent are common from year to year (i e , from one month to the same month a year earlier).

Estimation procedure The new estimation procedure makes use of what is referred to as a composite estimate The estimate for each item is a composite or weighted average of two estimates These two estimates of the same item are not independent but when properly weighted may yield a composite estimate with a smaller variance than either of the component estimates

(a) The first of these estimates, referred to as the regular ratio estimate, is obtained by essentially the same two-stage ratio estimating procedure that was used in making estimates from the 68-area sample The first-stage ratio estimates take advantage of 1950 Census information for counties The second-stage ratio estimates take advantage of current figures on the age-sex-color distribution of the civilian population of the United States In effect, after the sample returns have been multiplied by the first-stage ratio estimate factors, they are used to estimate the percentage distribution by employment status within an age-sex-color group, the percentages are then applied to the known total population for that group Thus the problem of estimating an aggregate (such as the total number of persons in the labor force) is

reduced to one of estimating percentages, with a consequent reduction in the sampling error for most of the items.

(b) The second estimate that enters into the composite estimate is the estimate for the preceding month to which has been added an estimate of the change in each item from the preceding month to the present month. The estimates from which the change is computed are made with the estimating procedure described in (a), but they involve for each month only the returns from the sample segments that are common to the 2 months (constituting 75 per cent of the sample).

The weights for the two components of such a composite estimate need not necessarily be equal, and for estimating any particular item optimum weights might be chosen. In this instance the weights used are equal, because equal weights in this case satisfy the condition that for most items there will be some gain in reliability of the estimates over those obtained from the procedure described in (a). The composite estimate takes advantage of accumulated information from earlier samples as well as the information from the current sample, and results in smaller variances of estimates of both level and change for most items, but the larger gains are achieved, for the most part, in the estimates of change.

The way in which the previous samples contribute to the current estimate can be seen if we write the composite estimate as:

$$x''_u = K(x''_{u-1} + x'_{u,u-1} - x'_{u-1,u}) + (1 - K)x'_u \tag{1}$$

where

$$0 \leqq K \leqq 1$$

x''_u is the composite estimate for month u

x'_u is the regular ratio estimate based on the entire sample for month u,

$x'_{u,\,u-1}$ is the regular ratio estimate for month u but made from the returns from the segments that are included in the sample in both months u and $u-1$, and

$x'_{u-1,\,u}$ is the regular ratio estimate for the previous month $(u-1)$ but made from the returns from the segments that are included in the sample in both months u and $u-1$.

Now, if we let

$$y'_u = K(x'_{u,u-1} - x'_{u-1,u}) + (1 - K)x'_u \tag{2}$$

$$x''_u = y'_u + Kx''_{u-1} \tag{3}$$

$$= y'_u + Ky'_{u-1} + K^2y'_{u-2} + K^3y'_{u-3} + \cdots$$
$$+ K^{u-1}y'_1 + K^uy'_0 \tag{4}$$

where y_0' is defined as the initial estimate for the month preceding the month for which the composite estimate is first made.

The variance of this composite estimate (under some simplifying assumptions) is given in the appendix. The between-p.s.u. variance is the same for the composite estimate as for the regular ratio estimate, because the same sample of primary units is used every month. The reduction in variance through the use of the composite estimate is entirely in the within-p.s.u. contribution to the variance. Because for many items the within-p.s.u. variance accounts for most of the total variance of month-to-month change, the gains on estimates of change are particularly worthwhile. For the rotation system described above and with some simplifying assumptions, Table 3 gives a few illustrative cases of changes in the within-p.s.u. variance when the composite estimate is applied.

TABLE 3

RATIO OF WITHIN-PSU VARIANCE OF COMPOSITE ESTIMATE TO THAT OF REGULAR RATIO ESTIMATE FOR SOME ALTERNATIVE ASSUMED CORRELATIONS*

Values Of ρ†			Ratio With $K=.5$‡		Ratio With $K=.6$‡	
ρ_1	ρ_2	ρ_3	Level	Month-to-month change	Level	Month-to-month change
0.95	0.90	0.85	0.73	0.44	0.66	0.36
0.90	0.85	0.80	0.76	0.54	0.72	0.49
0.70	0.65	0.60	0.87	0.80	0.87	0.79
0.50	0.45	0.40	0.99	0.94	1.04	0.95
0.00	0.00	0.00	1.26	1.10	1.44	1.14

* Assumes rotation of segments in sample with a segment in the sample four months, out eight months, in again four months, as described earlier. The correlations between estimates 11, 12, and 13 months apart are not given because they have a trivial effect on these estimates.

† ρ_1 is the correlation one month apart of estimates from identical segments,
ρ_2 is the correlation two months apart of estimates from identical segments,
ρ_3 is the correlation three months apart of estimates from identical segments.

‡ K is defined in Equation (1)

Effect of length of time in sample on response. One interesting point that affects the estimation method deserves further attention. If there were no measurement errors involved, the value of any characteristic measured from the sample would be exactly the same, on the average, whether measured from a rotation group in the sample for the first time, or from one that had been in the sample for one or more earlier months. This appears to be the case for most characteristics measured in the

labor force survey. For unemployment, however, there is evidence that households tend to report a somewhat higher unemployment rate the first month they are in the sample than in subsequent months. During 1948–1952, the difference between unemployment estimated from these segments in the sample for the first month and the average estimate from those segments in the sample for the five subsequent months was 0.2 per cent of the population 14 years of age or over This difference did not appear to vary with differing levels of unemployment, and reflected about a 10-per cent difference between the average unemployment rate for the rotation groups in the sample for the first time and the average for the five subsequent times (during a period when each rotation group was in the sample six successive months instead of following the rotation pattern described above)

Because the different rotation groups enter into the regular ratio estimate with different weights than they enter into the composite estimate, such differences for rotation groups mean that the expected values of the regular ratio and the composite estimates differ slightly. The average unemployment based on the composite estimate for the period 1948–1952 was 2 04 million compared to 2 09 million from the regular ratio estimate, a difference of about $2\frac{1}{2}$ per cent The average difference is smaller than the sampling error of either estimate for an individual month. Nevertheless, if we can ascertain why such a difference occurs, it should help in understanding and controlling response errors, this difference is the subject of investigation at the present time An explanation has not been found from the initial investigations.

Variance estimates As has been indicated, the variance estimates made from the 68-area sample were subject to relatively large variances. The 230-area sample provides much more reliable estimates of sampling variability than were possible with the 68-area sample. Because of the unreliability of the variance estimates made directly from the 68-area sample, a curve-fitting procedure was used to reduce the variance of the estimated variances Also, the curve-fitting procedure saved computing time and facilitated presentation of sampling errors While such methods will reduce the sampling variability of the variance estimate, they may yield biased estimates. The improved variance estimates from the 230-area sample suggest that some of the variances estimated for the 68-area survey were substantial underestimates, some others were substantial overestimates.

Until the acquisition of a high-speed electronic computer, the Univac, extensive approximations were introduced into the estimates of the variances to avoid computations that would be exceedingly time-

consuming with the available equipment. The availability of the Univac makes it possible to avoid most of these approximations. Even with the electronic computer, however, the work of making variance computations would be extremely heavy if variances were computed for all items directly Approximate methods will continue to be used in the future, but they will be evaluated by more exact computations than have been feasible in the past.

Quality Control. A system of statistical quality control on field interviewing was introduced as a part of the new survey design to help insure results of consistently acceptable quality from the CPS At the present time the quality control program is based on reinterviews of a subsample of the work of about 12 per cent of the enumerators each month. The principles applied are those of process control as described in the statistical quality control literature. The program has been designed to identify those enumerators whose work is beyond acceptable limits of performance so that they may be singled out for retraining or other administrative action The basic philosophy is to concentrate the principal supervisory emphasis at those points where it will be most effective—and to give less attention to interviewers whose work is under satisfactory control This approach is regarded as effective because of the relative stability of the interviewer force

Response errors Table 4 gives a frequency distribution of the average gross differences in results between original interviews and reinterviews covering six recent months of experience with the quality check. Results are presented for both "coverage," which refers to the enumeration of the population in the sample, and "content," which refers to the characteristics of those persons. The check results for each interview are used as the standard from which percentage differences or gross "error rates" are computed These differences between the original interview and the check are defined as "errors" although often they represent simply response variation without a necessary implication of an error having been made by the original interviewer.

In addition to furnishing information on individual enumerators, the reinterviews taken for the quality control operation provide a continuing source of information on the average overall quality of the CPS program. Table 5 shows results of the original and the check interviews over a 6-month period; it also shows the proportion of persons identified in a particular class in the reinterview who were identified in that same class in the original interview. Some of the items show relatively large differences between the original and check reports This is especially true for persons working part-time, those who have a job but

TABLE 4

DISTRIBUTION OF INTERVIEWERS' GROSS "ERROR" RATES* CPS CHECK, MAY–OCTOBER, 1954

Gross Error Rate (Per Cent)	Number of Inter-viewers	Cumulative Per Cent Of Inter-viewers	Number of Errors	Cumulative Per Cent Of Errors
Coverage: Total	229		224	
0	178	100	—	
0.1– 0 9	—		—	
1 0– 1.9	2	22	2	100
2 0– 2 9	—	21	—	99
3.0– 3.9	3	21	4	99
4 0– 4 9	7	20	8	97
5 0– 5.9	5	17	5	94
6 0– 6 9	2	15	4	92
7 0– 7.9	6	14	11	90
8 0– 8.9	6	11	21	85
9.0– 9 9	1	9	2	75
10.0–14 9	4	8	10	75
15.0–24 9	6	7	36	70
25.0 and over	9	4	121	54
Content: Total	232		1319	
0	72	100	—	
0.1– 0.9	27	69	34	100
1 0– 1 9	24	57	69	97
2 0– 2 9	29	47	159	92
3 0– 3 9	18	34	132	80
4 0– 4 9	14	27	128	70
5 0– 5.9	12	21	129	60
6 0– 6 9	14	16	180	51
7 0– 7 9	6	10	69	37
8 0– 8 9	2	7	41	32
9.0– 9.9	3	6	47	29
10 0–14 9	6	5	127	25
15 0–24.9	3	2	108	16
25.0 and over	2	1	96	7

* The check interview is carried out by a supervisor or special interviewer without knowledge of the results of the original interview. However, after each check interview is completed the results of the original interview are compared with the check results, and differences are discussed with the respondent in an effort to explain discrepancies and to obtain the most accurate responses feasible. Any corrections to the original check interview results are recorded. After this reconciliation, the original check interview results tend to be sustained for about two-thirds of the original differences and the original interviewers results are sustained for about one-third of the original differences. The results presented as "errors" in this table are based on the differences after such reconciliation.

TABLE 5

MONTHLY AVERAGE RESULTS FROM ORIGINAL AND CHECK INTERVIEWS* FOR PERSONS INCLUDED IN REINTERVIEW SAMPLE, MAY–OCTOBER 1954

Employment Status	Monthly Average Number Of Persons Based On*		Per Cent Identically Reported†
	Original interview	Check interview	
Labor force	1,221	1,242	93 1
Employed	1,149	1,168	93 5
Agriculture	158	164	94 5
Nonagriculture	991	1,004	93.3
Full time (worked 35 hours or more)	666	663	96.5
Part time (worked less than 35 hours)	253	266	86.9
With a job but not at work	72	75	86.7
Unemployed	72	74	86.5
Not in labor force	945	924	98 7

* Reinterview averages are based on results after reconciliation. See footnote *, Table 4.

† Per cent of persons reported in specified class in reinterview who were reported in that identical class in the original interview.

are not at work, and those unemployed, and thus the differences tend to be concentrated among those groups with marginal attachments to the labor force. Table 6 permits comparisons of average results for segments in the sample for the first time, and for continuing segments. Only a few of the differences between original and check results in Tables 5 and 6 are statistically significant.

Problems in measuring response errors and in controlling the quality of field interviewing vary directly with the difficulty in separating and measuring errors arising from limitations of the data-collection instrument (questionnaire, definitions, etc) itself, errors from the respondent, and errors arising from interviewer performance. In measuring characteristics, as distinguished from coverage, the response variability is sometimes large even when the interviewer carries through the questioning procedure exactly as specified. For such characteristics the minimum aim will be to attain a small net error (bias), but there is no assurance that efforts which are successful in maintaining a small net

error will result also in a small gross error. Thus, for some of the more difficult labor force and related measurements, the contribution of interviewer error to the gross error rate may be relatively small compared with those that are a consequence of the concepts and definitions employed and errors made by respondents. In such circumstances modification of the concepts and definitions may be needed.[6]

Problems in shifting from old to new sample. Funds which could be used for shifting from the 68 to the 230-area sample were not available until August 1953, when the decision was made to introduce the new sample during the latter part of 1953 and early in 1954. The survey work was initiated earlier in a few areas to allow opportunity for preliminary testing and modification of procedures Then, in November 1953 the survey work was started in one-third of the new areas; two-thirds were in operation in December; and the entire sample was in operation in January 1954 on a "dry-run" basis. The results of the old and new surveys for January were reasonably consistent for most items, but important differences were observed for some. For unemployment, in particular, the new estimate exceeded the old by about 30 per cent, a relatively large difference and greater than could have been explained by sampling variability. However, from some points of view the difference may not be large, even though statistically significant, for an item whose measurement often involves the evaluation of attitudes; one survey gave an estimate of a little less than 4 per cent of the labor force as unemployed, and the other approximately 5 per cent.

The use of a probability sample made it possible to estimate the differences that might reasonably occur from the change in samples. Because the observed difference for unemployment was larger than could be explained as sampling error, it was clear either that the sampling operations had not been carried out in accordance with the specifications, or that differences in measurement or response must account for an important part of this difference between the two survey results Failure to find important departures from the sampling specifications led to the inference that an important part of the difference results from other factors

[6] An illustration of a problem of this type may be found in the U S Census of Agriculture. Prior to the 1950 Census of Agriculture, the definition of a "farm" for purposes of the census enumeration depended for places of less than 3 acres upon the value of products sold or consumed by the operator's household. The enumeration of the value of products in those cases was found to be subject to relatively large response errors Some improvement in this case resulted from dropping dependence on value of products from the definition of a farm for the census enumeration and substituting a list of specific characteristics.

TABLE 6

MONTHLY AVERAGE RESULTS FROM ORIGINAL AND CHECK INTERVIEWS* FOR PERSONS INCLUDED IN REINTERVIEW SAMPLE, BY NEW AND CONTINUING SEGMENTS, MAY–OCTOBER 1954

Employment Status	Continuing Segments		New Segments	
	Original interview	Check interview	Original interview	Check interview
Total persons 14 years old and over	1,667	1,667	499	499
Labor force	942	957	279	285
Employed	889	901	260	267
Agriculture	128	131	30	33
Nonagriculture	761	770	230	234
Full time (worked 35 hours or more)	509	504	157	159
Part time (worked less than 35 hours)	200	210	53	56
With a job but not at work	52	56	20	19
Unemployed	53	56	19	18
Not in labor force	725	710	220	214

* Check averages are based on results after reconciliation See footnote *, Table 4.

One might expect that new interviewers recently recruited and trained in the measurement of the rather difficult concepts involved might be the principal source of difference, and that as these new interviewers acquired more experience the results from the new survey would approximate those of the old Undoubtedly this was a factor, but it appeared not to be the principal one. On the contrary, evidence indicated that the old sample, more than the new, was yielding results that were out of line For example, the new sample was providing higher estimates of some of the marginal classes of the unemployed and of other labor force categories, and past experience had indicated that more careful work was needed to identify more of these marginal cases. Also, during the last few months of 1953 and especially in January 1954, the number of persons receiving unemployment insurance benefits had become an increasingly higher fraction of total unemployment as estimated from the 68-area survey, and a considerably higher fraction than had been observed at most dates in earlier years

One factor that might explain the relationships appeared to be par-

ticularly important. During the period of introducing the new sample in the field, beginning in September 1953, the supervisors had been instructed to concentrate their attention on recruiting and training the staff for the new survey, and on getting the new sample into operation. The experienced interviewers were to carry on for several months with training and review of their work by mail This was a sharp change in procedure—this staff had been accustomed to continued personal attention. At the same time, most of the old interviewers probably were hearing rumors that their jobs would be terminated shortly, even before they were officially notified.

This type of reasoning suggested that the performance of the experienced interviewers might have deteriorated, and that it would be desirable to reinforce the supervision and training activities in the old sample as well as the new before the February survey Accordingly, an additional special training session was arranged in February for the interviewers in both the 68-area and 230-area designs.

The results in February came considerably closer together—sufficiently close that the differences might be explained as entirely due to sampling variability. There was, however, some supplementary evidence that other factors might still be present, although in a considerably lesser degree than in January. It appeared that because of the retraining, or perhaps as a result of the widespread publicity regarding the difference in the unemployment estimates and the effect this publicity might have had on the interviewers, or perhaps because of both of these and other factors, the differences were largely removed in the February survey. The 68-area sample was discontinued in March.

Table 7 shows the estimates of the major labor force characteristics for January and February from both samples, together with the sampling errors of the differences between the estimates.

FUTURE DEVELOPMENTS

When the new survey results were published in January, the Secretary of Commerce appointed a Committee[7] to review the reasons for the differences in the new and old survey results. This Committee and the Census staff both have worked extensively in exploring the results obtained and have outlined areas for further work A continuing research program has been in progress on a small scale, and funds are being requested for expanding this research Topics for particular attention include further work on response errors in labor force measure-

[7] Consisting of Frederick F Stephan, Chairman, Lester R. Frankel, and Lazare Teper.

TABLE 7

PERSONS IN CIVILIAN LABOR FORCE BY EMPLOYMENT STATUS, AS ESTIMATED FROM 230-AREA AND 68-AREA SAMPLES, WITH SAMPLING ERRORS OF DIFFERENCES, JANUARY AND FEBRUARY 1954

Employment Status	230-Area Sample (1000's)	68-Area Sample (1000's)	Difference (d) (1000's)	Estimated Standard Error Of Difference (s) (1000's)	$\frac{d}{s}$
January					
Civilian labor force	62,840	62,137	703	591	1 19
Nonagricultural employment	54,469	54,433	36	756	0.05
Agricultural employment	5,284	5,345	−61	353	−0 17
Unemployed	3,087	2,359	728	188	3.87
February*					
Civilian labor force	64,079	63,491	588	603	0.98
Nonagricultural employment	54,535	54,480	55	757	0.07
Agricultural employment	5,761	5,626	135	376	0.36
Unemployed	3,780	3,385	395	233	1 70

* The February estimates for the 230-area sample given here differ from the published estimates which are based on a composite estimating method The estimates given here are based on the regular ratio estimate using the February sample and are the appropriate estimates for this comparison.

ment, questionnaire design, interviewing, the effectiveness of alternative training procedures, and quality control methods Attention is being given to identifying the reasons for the small but persistent difference in unemployment measurements for households in the sample for the first time, and the measurements obtained from interviews in the same households in subsequent months.

In an attempt to realize the objective of maximum information per dollar spent, a particularly important class of problems arises out of questions as to the effects of training, observing, checking and other "non-productive" work on the quality of data obtained. Objective evaluation of these effects is urgently needed to replace the present intuitive guides which may often be misleading We hope to arrive at

this evaluation by experiment and test The existence of such data may have far-reaching effects on the organization and administration of field surveys.

There is a need, also, for careful study of accuracy requirements in the results for various purposes. This problem is exceedingly difficult and long neglected; it needs extensive attention of the Census staff and of the users of the labor force statistics

APPENDIX—WITHIN-P.S U. VARIANCE OF THE COMPOSITE ESTIMATE

[Prepared by Max Bershad and Margaret Gurney]

A *Monthly Level*

From equation (4)

$$x''_u = \sum_{i=0} K^i y'_{u-i} \tag{4a}$$

Each y'_u can be expressed as the expected value plus a random deviation resulting from the selection of p.s.u.'s and a random deviation resulting from the selection made within the selected p.s.u 's For a fixed selection of p s u.'s, the conditional expected value, over all possible within samples, of the product of the within deviation and the between deviation is zero; and consequently the expected value of the product is zero when taken over all possible selections of p s.u.'s This is also true when the within deviation for one month is multiplied by the between deviation of another month. The two types of deviations are uncorrelated and the total variance is expressible as the sum of the between-p.s u variance and the within-p.s u variance.

The sample of p s u 's is unchanged from month to month and the between-p s u variance of the composite estimate is unchanged from that of the regular ratio estimate

The within-p s u. variance of the composite estimate is derived below, where it will be understood that the symbols for variance and covariance refer only to the within-p.s u component

Consequently from (4a)

$$\sigma^2_{x''_u} = \sum_{i=0} [K^{2i}\sigma^2_{y'_{u-i}} + 2K^{1+2i}\sigma_{y'_{u-i}y'_{u-i-1}} + 2K^{2+2i}\sigma_{y'_{u-i}y'_{u-i-2}} + 2K^{3+2i}\sigma_{y'_{u-i}y'_{u-i-3}} + \cdots] \tag{5}$$

or

$$\sigma^2_{x''_u} = \sum_{i=0} K^{2i}\sigma^2_{y'_{u-i}} + 2\sum_{i=0} K^{2i} \sum_{j=1} K^j \sigma_{y'_{u-i}y'_{u-i-j}} \tag{5a}$$

If

$$\sigma^2_{y'_{u-i}} = \sigma^2_{y'_u}$$

for all i and if,

$$\sigma_{y'_{u-i}y'_{u-i-j}} = \sigma_{y'_u y'_{u-j}}$$

for all i and j, then (5a) becomes, for large u,

$$\sigma^2_{x''_u} = \frac{\sigma^2_{y'_u}}{1-K^2} + \frac{2}{1-K^2}\sum_{j=1} K^j \sigma_{y'_u y'_{u-j}} \tag{6}$$

B. *Monthly Difference*

Since

$$x''_u = y'_u + Kx''_{u-1}$$

and

$$x''_{u-1} = y'_{u-1} + Kx''_{u-2}$$

then

$$(x''_u - x''_{u-1}) = (y'_u - y'_{u-1}) + K(x''_{u-1} - x''_{u-2}) \tag{7}$$

Letting

$$z''_u = x''_u - x''_{u-1}$$

and

$$w'_u = y'_u - y'_{u-1}$$

we have from (7)

$$z''_u = w'_u + K(z''_{u-1}) \tag{8}$$

which is of the same form as the composite estimate for monthly level. Thus the variance of the month-to-month difference is given by (6), with y'_u in (6) being replaced by $y'_u - y'_{u-1}$; and

$$\sigma^2_{z''_u} = \frac{\sigma^2_{w'_u}}{1-K^2} + \frac{2}{1-K^2}\sum_{j=1} K^j \sigma_{w'_u w'_{u-j}}$$

SAMPLING METHODS IN THE YUGOSLAV 1953 CENSUS OF POPULATION*

S. S. Žarković

Yugoslavia Federal Statistical Office

PROGRAM

During the preparations for the 1953 census of population in Yugoslavia two groups of problems appeared, the solution of which required the application of sampling methods. In the first group were problems common to all population censuses; the second contained a question of special interest to less developed countries.

One of the problems in the first group related to the time required for processing the data. The censuses of 1921, 1931, and 1948 were tabulated by hand. The processing of the 1921 census lasted 10 years. The 1931 census was still incomplete at the beginning of the war in 1941, when all of the material was lost The processing of the March 1948 census lasted to the end of 1949 in spite of its narrow scope and the great number of people engaged in it.

Results so delayed lose much of their practical value. For the 1953 census it was essential that estimates of the most important facts be prepared as quickly as possible, since social and economic measures depended on census findings. The first problem therefore, was to prepare a set of preliminary estimates that could be used until better and more complete figures became available.

As will be shown later, the application of sampling methods in this census was such that the preparation of advance estimates could be divided into two parts, (a) estimates prepared very quickly for a small number of characteristics, and (b) estimates prepared over a somewhat longer period for a larger number of characteristics.

Another of the common group of problems is the completeness of enumeration of the population Investigation of this point was important for several reasons. First of all, a measure was required of the accuracy of the population totals obtained by the enumerators for 1953. Then, some statistical questions relating to the next census also called for putting this check on the program. Checking completeness shows in what cases enumeration meets difficulties and what problems, in the given system of enumeration, should be considered more fully

* The author is very much indebted to Morris H. Hansen, U. S. Bureau of the Census, who read the manuscript and made a great number of suggestions for improving it.

in future work. It gives information on the capabilities of the available enumerators, shows the advantages and deficiencies of the definitions used, of the accepted system of enumeration, etc.

The next problem is the measurement of the volume of errors in answers on census questions. This is a modern theme and a very important one for the proper formulation of questions, for the organization of the census, and for the users of statistical data as well. The general importance of the problem of errors in answers is clear enough, but such research has a special value in countries having a relatively low level of cultural development, with a large percentage of illiterate people, who are not familiar with the concepts used in the census. Measuring the volume of errors under such conditions makes it possible to see what can be done by the census and what reliance can be placed on the data obtained The usefulness of such research is much greater if it is designed in such a way as to give the percentage of wrong answers on a question and information on the sources of errors as well. This information is valuable in contemplating measures for reducing errors or for improving answers.

The last problem in this group is the quality of the editing Although a great many errors appearing on census questionnaires are caught in the process of editing, by no means all are eliminated. And it is known that many of those remaining on the questionnaires could in principle have been corrected. It was important, therefore, to determine to what extent errors of different types were corrected in editing in this census, and, more generally, to what degree the editing approached the ideal in quality of work. Without answers to these questions it is not possible to gain full insight into the reliability of the census results or to prepare adequate editing procedures for the next census.

In the second group of problems—those of special concern to less developed countries—we undertook a study of the accuracy and value of literacy data. As explained in a previous article published in this *Journal* [12], the first impetus for doing so came from the strange course of the figures on percentage of illiterates as shown in the previous censuses. These results were as follows:

Census	*Percentage of Illiterates*
1921	50.5
1931	44.6
1948	25 4
1953	24.9[1]

[1] This figure represents the sample estimate.

The sharp drop in illiteracy from 1931 to 1948 (averaging over one per cent a year) and the negligible decrease from 1948 to 1953 seem clearly open to question, since the former period included the war years, when most of the schools were closed, while the latter period was characterized by an extensive expansion in the country's educational system and by special efforts to eliminate illiteracy.

In addition to this uncertainty as to the reliability of the census figures, there was need for further light on the meaning of the replies given regarding literacy. Within the group reporting themselves as able to read and write, what degrees of proficiency were represented? How many were fully literate, how many moderately so, and how many were on the borderline of complete illiteracy?

In summary, then, the application of sampling methods in this census covered the following program:

(i) first advance estimates,
(ii) second advance estimates,
(iii) checking the completeness of enumeration,
(iv) investigation of the volume of errors in answers,
(v) checking the value of literacy data, and
(vi) checking the quality of editing.

First, we shall give some data on the samples used in carrying through this program and then we shall indicate briefly what was done and what results were obtained with each particular problem.

THE GENERAL SAMPLE

For all phases of the program, enumeration districts were used as the primary sampling units. For parts (ii) and (vi) special samples were drawn, but for the other four parts of the work the same primary sample was employed Information regarding this general sample will therefore be given here. The samples for parts (ii) and (vi) will be described later.

Ten days before the beginning of the census the total number of enumeration districts in each administrative unit (commune, county, republic) was known centrally. On the basis of the administrative division all the enumeration districts were stratified into two strata: rural and urban From the urban stratum 100 enumeration districts were selected and from the rural stratum 149. The selection of enumeration districts as primary units was done with equal probabilities but with a smaller sampling fraction in the rural stratum than in the urban.

In determining the size of the sample of primary units two considerations were decisive. The first was the available number of inspectors. The study program was organized in such a way that only one in-

spector was assigned to each enumeration district and there he had to do a rather complicated job. Therefore the inspectors' training needed to be intensive But due to conditions under which the surveys were prepared the number of qualified inspectors was limited and so the sample of primary units had to be less than 300. The second consideration related to point (iii) of the program. It was desired that the measurement of the completeness of enumeration should give an estimate of the total population with a coefficient of variation that would not exceed 0 30 per cent On the basis of advance speculations we expected the coefficient of correlation between enumeration districts of the number of residents as obtained by the census and the same number as obtained by the check to be as high as 0 995 It was possible, therefore, to satisfy the above requirement with a sample of 250 enumeration districts.

The sample of primary units was used without subsampling to get answers on points (i) and (iii) of the program. For points (iv) and (v) it was necessary to subsample individuals from the selected enumeration districts For each of these parts of the program the second stage sample was different, since the populations of individuals covered in the two studies were not identical. Further information on these samples will be given in subsequent sections

Table 1 contains information on resident population and on the number of enumeration districts in the population and in the sample, classified by strata More details on this sample are given in [1].

TABLE 1

BASIC DATA ON POPULATION AND SAMPLE

	Urban Stratum	Rural Stratum
Resident population (in thousands)	4,832	12,095
Total number of enumeration districts	29,805	89,194
Number of enumeration districts in the sample	100	149
Per cent of enumeration districts in sample	30	.15
Per cent of persons in sample	35	.16

All of the data in this table are based on the preliminary count.

FIRST ADVANCE ESTIMATES

The inspectors employed in this program came to their enumeration districts eight to ten days after the census day At this time enumeration was over and for each person in the country a filled questionnaire existed (theoretically). To get the first advance estimates it was initially

planned that the inspectors should take with them all questionnaires for the persons enumerated in their districts. The processing of these questionnaires would make possible the desired estimates. This system would give results without any duplication in processing because the punch cards for the sample districts can be put later in their proper place among the others.

The urgent need for data changed this plan in the following way: answers of all individuals enumerated in a sample district were summarized on a sheet by the inspector assigned to that district, and the totals so obtained were used to get the estimates. The census day was March 31 and these first estimates were ready at the end of April. So the basic results of the census were known immediately after the enumeration was completed. The following groups of tables were included in these estimates: (i) demographic data, (ii) nationality, (iii) literacy, (iv) degree of education, (v) religion, and (vi) economically active population.

From the point of view of variances the system of estimation used was not efficient as compared with a simple random sample of individuals, since some characteristics, such as branches of industry, occupational groups, and in general all economic characteristics, have a relatively high intraclass correlation within enumeration districts. For comparison we give the following data. The relative standard errors of the proportions of literate persons, as obtained by using the ratio method of estimation with enumeration districts as sampling units, were 2.5 per cent in the rural stratum and 1 9 per cent in the urban stratum. If the equal number of individuals were selected at random, the variances of the corresponding simple unbiased estimates would be 0.32 and 0.27 per cent respectively The difference in efficiency is considerable, but since obtaining these estimates was inexpensive the system will probably be retained in the future, the only change being that the number of primary units will be increased. It is easy to use the larger sample of primary units for this purpose because the compiling of the necessary data is simple and does not require the presence of a trained inspector

In similar cases we are usually interested in subsampling enumeration districts. In this work the subsampling was not used since it was found it would not decrease the costs.

SECOND ADVANCE ESTIMATES

The sample of 249 enumeration districts was not sufficient to yield satisfactory estimates in those cases in which a cross classification was

necessary. It also was inadequate for preparing separate estimates for the six republics. Thus there appeared to be need for another sample, larger in size. This second sample was desirable also because of the fact that the individual data used in preparing the first estimates were not previously submitted to editing, and because the regular processing of the census was expected to be finished toward the end of 1955.

The second sample, designed to meet the need for more detailed figures, also consisted of enumeration districts as primary sampling units; these were selected with a different sampling fraction for each republic, the fractions varying between 0.14 and 0.67. In the second stage of selection 10 per cent of households were taken. The use of different sampling fractions in the first stage selection resulted from the desire to have the sampling errors on pretty much the same level in each republic.

The relative efficiency of the household as the second stage sampling unit in comparison to the simple random sample of individuals is also low for some characteristics. An experiment was carried out in one republic to compare the relative efficiencies of different sampling units and different methods of estimation. The results of this experiment are summarized in Table 2 for some selected items, with the variance of a simple random sample of persons taken as 100 The stratification was done according to size of household.

On the basis of such data and after having taken respective costs into consideration, the household was selected as the more economical secondary unit.

The procedures used in connection with this sample, the results obtained, and the discussion of problems faced are presented in references [2, 4, 6, and 8]. The latter two publications also contain some tables showing the magnitude of standard errors broken down by items, republics and provinces.

COMPLETENESS OF ENUMERATION

For evaluating the completeness of census enumeration the general sample of 249 primary units was used, without subsampling, according to the following plan.

After the enumeration was completed, inspectors were sent to the selected enumeration districts provided with maps on which the borders of the districts were drawn. Their duty was to visit all dwelling units and make a new count of all persons, entering only their names and classifying them according to three groups: (i) permanently present, (ii) temporarily absent, and (iii) temporarily present. The first two

TABLE 2

RELATIVE EFFICIENCIES OF DIFFERENT SAMPLING UNITS AND DIFFERENT METHODS OF ESTIMATION*

Item	Sampling Unit				
	Person		Household		
	Simple random sample	Stratified sample	Simple random sample	Stratified sample	Ratio method
Males	100	100	70	147	143
Illiterates	100	100	71	71	71
Persons by year of birth					
1940–1948	100	108	59	87	103
1900–1909	100	104	90	100	96
1879	100	110	91	91	87
Employed persons					
Total	100	102	52	52	71
Workers	100	100	66	73	67
Employees	100	100	80	81	69
Farmers	100	100	30	42	40
Others	100	100	92	92	85

* The figures in this table are published in somewhat different form in Blejec [4]

groups constitute the resident population. In this new count the inspectors did not know the results obtained in the census Their task was to reproduce the situation on the census day.

After their count was finished the local population census commission (a group of local inhabitants charged with supervising the census operations in the field) matched the results of the checking against the results of the census Lists of the names made in checking by households were used to facilitate this matching. If differences appeared in any case they were subjected to a new check by the commission The results obtained in this new check were final Thus data obtained by inspectors were not considered as definitely correct This practice was justified later by the errors found in inspectors work In some cases they omitted to count a person and in still more cases their classification of people under the three above groups was not correct

This matching of the two sets of results serves also to eliminate errors that might appear because of the fact that the inspectors necessarily make their check of the completeness of enumeration at least a week

after the census day This time interval between the two surveys could be reduced by the simultaneous work of both enumerators and inspectors, but this would not give the desired result, since the enumerator would be informed about the check on his district and the whole purpose of checking would be lost. Thus a difference of several days is necessary. But if inspectors, because of population changes, were not always successful in reproducing the situation on the census day, the above matching and the following check should reduce their errors

On the basis of the results of this survey an indication of the quality of enumerators' work can be arrived at in the following way All persons in the selected enumeration districts were classified in one of the following classes

(i) persons not classified by enumerator (sign $-$),

(ii) persons given the same classification by both enumerator and inspector (sign $=$),

(iii) persons enumerated in the two surveys but differently classified (sign $\neq$),

(iv) persons classified by enumerator but not by inspector (sign $+$).

The percentages of cases falling in these classes are presented in Table 3 They show the different aspects of the quality of enumerators' work

As regards the number of persons in the resident population, the figures in Table 4 are a rough estimate

It will be seen that while a net under-enumeration was found in the urban stratum the rural stratum and the total population showed a net overenumeration. This result is somewhat unusual Hansen, Hurwitz, and Pritzker, [7] have shown an over-all net underenumeration in a similar check carried out in connection with the 1950 U S.

TABLE 3

PERCENTAGES OF CASES FALLING IN THE FOUR CLASSES

Class	Percentage	
	Rural Stratum	Urban Stratum
$-$	0 23	0 83
$+$	0 26	0.45
$\neq$	0 58	1 38
$=$	98 93	97 34

TABLE 4

ESTIMATES OF THE COMPLETENESS OF ENUMERATION OF THE RESIDENT POPULATION

	Stratum		Total	
	Rural	Urban	Number	Percentage
Under enumeration	27,500	21,500	49,000	0.29
Over enumeration	40,000	11,500	51,500	0.30
Gross error	67,500	33,000	100,500	0.59
Net error	+12,500	−10,000	+2,500	0 01

census of population. This seems to be a logical expectation in statistical surveys. But our result is explained by the fact that many people, belonging to the rural stratum, have a job outside the place where their families live. This is probably because of the housing shortage. Some of these people come home once a week, and others only from time to time. Census instructions concerning such persons were complicated and the result was double enumeration, at home and in the place of work. Similar difficulties were also met in the case of students absent from their parent's homes to attend school.

Some comment should be added regarding the degree of completeness of enumeration. Judging by the results, the enumeration was excellent from the point of view of completeness. It goes without saying that this could arise either from a very good initial enumeration, or from an inadequate check, whatever the quality of the original enumeration. The full answer on what really happened here is almost beyond practical possibilities because it would require further superimposed checks. The only additional evidence in favor of the general character of our results is a recheck made on 21 enumeration districts; the findings were the same except for one person. Other and more subjective evidence appears in the reports on this point from the general field inspectors of census operations; they also talk about the good work of the regular enumerators

For the improvement of census methods and techniques in this sort of research, great importance attaches to the analysis of cases in which the census and the check disagree. It was found here that disagreement rarely appears because the enumerators failed to classify someone. Most of the differences relate to difficulties met in applying the definitions to individual cases. For illustration we might mention the examples of two enumeration districts. In the first district a tubercular

sanatorium was located, and in the second a building site for a new factory. Each of these districts had frequent population changes; some people would remain for a month, others for several months and even years. In these circumstances it is almost impossible for an enumerator to carry through the census definitions. Furthermore the definitions themselves were not always formulated in such a way as to fit the cases met in these two enumeration districts. Details concerning this check are given in [10].

ERRORS IN ANSWERS

Errors in answers on census questions may appear in connection with any enumerated person regardless of age or other personal characteristic and also regardless of whether the respondents filled out the questionnaire themselves (as was provided in instructions) or whether it was prepared by the enumerator (for illiterates, children, etc.) or by some third person (as actually happened in many cases). Accordingly, for the second stage sample used in this part of the research program, all individuals enumerated in the 249 selected primary units had to be taken into account In this case the sample of secondary units was drawn by inspectors on the field. As a frame they used the filled questionnaires and applied the sampling fraction of 1:8 in the rural stratum and 1:10 in the urban stratum. Thus in the rural stratum, a total of 2,470 individuals was selected and 1,684 persons in the urban stratum. Different sampling fractions in the two strata were used in order to get an approximately equal distribution of work among the inspectors

The fieldwork was carried out as follows Without being informed of the answers given previously in the census, inspectors got in touch with each person in the sample and took again the answers on the more important census questions. In doing so they used some control questions and asked for documents to prove the accuracy of answers obtained (provided documents were available and the respondents were willing to show them). Their results were then matched against the census results In this matching, identical answers were defined as correct answers; for most circumstances this may be considered as a sound supposition, but it does not necessarily imply an absolutely correct answer. For example, if someone, for one reason or another, gives a wrong answer on the question "Occupation" and later declares the same to the inspector, his answer in this check would be considered as correct in spite of the fact it is actually wrong. Here again the problem appears as to how far one can go in using different means of checking. For reasons of simplicity and economy our procedure was as described, intro-

ducing the corresponding limitation in the value of the results obtained.

As opposed to this, cases in which differences were found were subjected to a new check, made this time by more rigorous methods. Inspectors themselves and the members of the local population commissions were charged in such cases to look for documents if available at accessible places, to talk to other persons able to provide information about the individual in question, or to get again in touch with that individual, throwing more light on the problem by using additional questions The resulting answer was accepted as correct

For the final analysis, the following definition of error was used. each answer in the census is wrong if it is not identical with the corresponding answer in the check. In addition to this class of errors, formal errors were also taken into account An example of these errors is the following: If a person on the question "Degree of education" entered a dash, this was counted as an error, since the instructions explicitly required that answers be written in full. Thus the concept "Volume of errors" adopted in the analysis contained both classes of errors.

Considering first the questionnaire as a whole, and defining as correct those showing no differences with the check, and no formal errors on any questions, we find the results presented in Table 5, where the six republics are ranked according to the percentage of correct questionnaires It is interesting to note that these ranks correspond to the ranking of the republics according to the level of their cultural development as measured by the percentage of literate people and the degree of education. For the professional statistician this is important information.

TABLE 5

RANKS OF REPUBLICS ACCORDING TO PERCENTAGE OF CORRECT QUESTIONNAIRES

Republic	Percentage
Slovenia	59.1
Croatia	54 0
Serbia	45 9
Montenegro	40 9
Bosnia and Herzegovina	36 8
Macedonia	24.8

If we classify the questionnaires by the number of errors in them we get the percentages shown in column 1 of Table 6. Analyzing the

TABLE 6

PERCENTAGES OF QUESTIONNAIRES WITH A GIVEN NUMBER OF ERRORS

Number of Errors	Percentage of Questionnaires	
	Without editing	After editing
0	46 2	62 0
1	29 1	21 9
2	12 7	7 3
3	4 7	2.6
4	3 0	2 6
5	2.2	1 9
6	1 0	0 9
7	0 6	0 3

results by sex, it was found that the percentage of wrong answers for males is somewhat lower than the percentage for females

With regard to the percentage of wrong answers on particular census questions, the results obtained for the rural and the urban strata are presented in columns 1 and 2 of Table 7. These percentages are calculated with respect to the total number of persons in the sample This is why some of them are low This is the case particularly with questions which only a limited number of people answer (e g , the number of live births) It will be noted that there is a high rank correlation between the percentages of wrong answers in the two strata.

As a further step in the study, the possibility of correcting errors in answers by editing was considered The answers as finally determined by the check were known and hence it was possible to estimate what would happen to errors as a result of maximum editing. For this purpose three classes of errors need be distinguished. In the first class are errors that can be corrected in editing and that do not appear in tabulation An example of such an error occurs in determining age from the date of birth If the age is tabulated by years, those errors in dates of birth that do not change the age in full years disappear as errors during tabulation. The second class comprises errors that can be identified as inconsistencies but can be corrected only if some additional information from the field were obtained for the particular person. In the third class are those errors that can not be eliminated because there is no basis for doubt regarding the reliability of the answers given. We have such a case if someone falsely reports his date of birth

TABLE 7

PERCENTAGES OF WRONG ANSWERS ON CENSUS QUESTIONS

Question	Percentage of Wrong Answers			
	Without editing		After editing	
	Rural stratum	Urban stratum	Rural stratum	Urban stratum
	1	2	3	4
Date of birth	32.9	17.5	14 9	7.7
Place of birth	2.6	1.3	2.6	1 3
Marital status	2.3	4.0	1 0	1.4
Number of times married	2.8	4 2	1 1	1 7
Age at first marriage	13 4	12 4	13.4	12.0
Number of live births	1 6	1.5	1 5	1.3
Number of children on the census day	0 5	0 7	0 3	0.5
Legal nationality	0 1	0.4	0.1	0.4
Literacy	5 3	6.0	4.8	2.7
Degree of education	12 7	14.9	8.7	10.6
Occupation	11.0	11.3	5.3	6.9
Occupational status	7.8	5.3	7.0	3.7
Industry	7.0	8.5	6.3	5.4
Economic sector	6 8	6.8	6 0	4.5

by several years or enters the wrong occupation or the wrong degree of education.

In order to estimate the maximum effect of editing or the minimum percentage of errors that would necessarily remain in the tabulated census figures, we assume that the editing will be carried through ideally, i e , all errors of the first and the second class will be corrected. This assumption appears justified because of the fact that the statistical service in Yugoslavia has a very wide-spread field staff which was used in this census to get the additional information needed regarding some persons for correcting errors of the second class. The frequencies of errors in questionnaires after editing are shown in column 2 of Table 6; the percentages of errors that remained in answers to various census questions after editing are presented in columns 3 and 4 of Table 7. As one sees, editing may considerably reduce the number of errors but its influence is not equal for all questions.

The results of further analysis for some particular census questions might be noted.

In connection with answers on the question "Date of birth" the following was found:

(i) the percentage of wrong answers is larger in the rural stratum,

(ii) more correct answers are given by males than by females,

(iii) the rank of the republics according to the percentage of wrong answers on this question is the same as their rank according to the degree of cultural development,

(iv) no systematic tendency to increase or to lower the age was found among the wrong answers;

(v) the wrong dates of birth show the characteristic tendency of concentration about certain dates, primarily about those ending in 0 or 5. This holds for the two strata;

(vi) there are some dates with relatively low frequencies, such as 2, 3, 22, 23, 24, 27, 29, 30 and 31;

(vii) the frequency of births, according to wrong answers, is considerably larger in the first 15 days of months than in the remainder although in the latter there are more days (because of the 31st in some months).

With regard to the question "Age at first marriage" it was found that males had no tendency to change the real age in either direction. For females the results show a definite tendency to overstate this age (probably because some girls marry very young).

Understatement was found in wrong answers on the question "Number of live births."

For literacy no tendency was discovered either to over- or to understate the real status This contradicts to some extent the widespread opinion that a large proportion of illiterate persons report themselves as literate

With regard to the degree of education the same number of over- and understatements was found. The results also show that in the majority of cases the wrong answers change the degree of education for only one school year. In examining the sources of these errors an interesting result was found, i e., the same tendency to round off the degree of education as was found in reporting date of birth. This rounding off is made on some characteristic educational level such as the completion of elementary school, high school, or a full college education.

The analysis of wrong answers on the "Occupation" and "Industry" questions shows that a large proportion of the errors have their source in difficulties in determining the occupation of housewives, particularly in the rural stratum. This information will be very useful in preparing instructions and adequate definitions for the next census.

The problem of errors in this census, their analysis, the organization of this survey and some related questions are discussed in [9] and partially in [3].

THE VALUE OF LITERACY DATA

Not all persons in the general sample of primary units have been included in the study of the value of literacy data. Since the census questions regarding literacy were confined to persons 10 years of age and over, those under 10 years of age were automatically excluded. Furthermore, persons reported in the census as having completed elementary school or more were not considered, as it was assumed that they were definitely able to read and write. And on parallel grounds, those reported in the census as illiterate were excluded from the analysis. From the remaining individuals a secondary sample was drawn, covering 417 persons in the urban stratum and 1,022 in the rural stratum.

To investigate the degree of the ability to read and write, a series of special tests were prepared, with a system of points for right answers or successful achievement that allowed a maximum of 45 points for reading and ten points for writing These tests were given by the inspectors to each person included in the sample The results obtained show that the literacy of the "literate" population is a continuous variable, which may have values falling anywhere within the range of zero to the maximum number of points possible in the tests

This brings out clearly the complexity of the problem of literacy data. Individual statements on literacy are merely the reflection of personal opinions as to what literacy is It was found that persons may consider themselves literate when their ability to read and write is at such a low level that for any practical purpose they would have to be placed in the illiterate group. It is obvious that literacy statistics could be improved if one were able to give a quantitative definition of the degree of literacy required for an affirmative answer, but unfortunately this cannot be done in the census The only practical means of throwing more light on the meaning of census data on literacy appears to be through a supplementary sample survey such as reported here

Further information regarding this study will be found in [1, 12,13].

QUALITY OF WORK IN EDITING

The research of editing was undertaken to determine the quality of work in this phase of processing, the extent to which errors remained

that could have been corrected and the specific properties of these errors that were responsible for their being not corrected.

Perhaps the best solution of this problem would be the use of quality control methods as described in [5] in connection with the verification of punching This system, however, could not be used in this census because of the fact that the necessary professional personnel were engaged in other parts of the program. What was possible, was a study of editing on the basis of the corrections appearing on the census questionnaires after the editing was completed.

The experiments carried out so far show that this survey should be taken on a relatively large number of enumeration districts as primary units and on a small number of households as secondary units within each selected enumeration district. Since all editing of questionnaires for a particular enumeration district was handled by one person, the quality of work within enumeration districts was homogeneous. Furthermore, the households within enumeration districts also tend to be homogeneous with respect to many characteristics. A satisfactory insight into the quality of editing, therefore, requires a large sample of primary units and a small sample of households within each primary unit. This procedure is supported by the fact that in our conditions of storing census material the access to data on both enumeration districts and households is relatively easy, while this does not hold for data on individuals.

This analysis of the quality of editing is considered very important, since our study referred to earlier has shown that the classical way of editing answers on all questions could well be replaced by some more economical procedure

At the moment of writing this work is not yet completed

CONCLUSION

The most important fields for the application of sampling methods in connection with censuses are as follows: (i) pretesting, (ii) getting answers on certain additional questions, (iii) coverage check, (iv) evaluation of the volume of errors in answers, (v) quality control of the processing, (vi) advance estimates, (vii) tabulation, and (viii) replacement of the complete census by the sample census.

All of these points were not included in the work reported here, since the full program would have required a much longer period of preparation and experimentation than we had at our disposal. In our opinion, a systematic use of sampling methods in connection with a census of

population should not be started on a broad front until a certain amount of experience from the same or very similar work had become available. Before we started we did not know the answers to many questions of fundamental importance for such surveys. Will it be possible to find adequate inspectors? What will be the quality of their field work? How are the people going to react to our demands? What ways and resources will prove best in looking for proofs of the accuracy of census results? What is the magnitude of the components of the variances in different surveys? and so forth. Without having such basic information it would be hazardous to start with a large program. Therefore, the main emphasis in this work was put not on obtaining results of interest to the "statistical public" at large, but on securing information about how to proceed in the next census. In this connection a number of experiments are still going on at the moment of writing and still others are planned. But the results already obtained have clarified many problems important for the application of sampling methods to the points discussed above, and have also thrown light on many other dark places in the last census for which more adequate solutions will, we trust, be found in the next census.

BIBLIOGRAPHY

[1] Balaban, V; Ispitivanje pismenosti [An inquiry into the literacy data], *Statistička revija*, IV, 1954.

[2] Blejec, M., Introductory remarks to "*Vzorec prethodnih rezultatov popisa prebivalstva 31 marca 1953 v LR Sloveniji*" [Sample estimates of the results of the census of population as of March 31, 1953], Statistical Office, Ljubljana 1953.

[3] Blejec, M., Kontrola tačnosti zbranih podatkov popisa prebivalstva 31 marca 1953 v LR Sloveniji [Checking accuracy of the results of the population census as of March 31, 1953], Statistical Office, Ljubljana 1953.

[4] Blejec, M., Uzorak prethodnih rezultata popisa stanovništva u 1948 i 1953 godini u NR Sloveniji [Sample results of the census of population as of 1948 and 1953 in P.R of Slovenia], *Statistička revija*, III, 231–42.

[5] Daly, F. J., and Guilford, L.; "Sample verification and quality control methods in the 1950 census," in Hansen, M. H., Hurwitz, W. N., and Madow, W. G, *Sample Survey Methods and Theory*, New York 1953, Vol. I.

[6] —:*Ekonomska obeležja stanovništva, rezultati uzorka* [Economic characteristics of the population. sample results], *Statistical Bulletin* No. 28, Federal Statistical Office, Beograd 1954.

[7] Hansen, M. H., Hurwitz, W N., and Pritzker, L., "The accuracy of census results," *American Sociological Review*, 18 (1953), 416–23

[8] :*Vitalna, prosvetna i socijalna obeležja stanovništva; rezultati uzorka* [Vital, educational, and social characteristics of the population: sample results], *Statistical Bulletin* No. 29, Federal Statistical Office, Beograd 1954.

[9] Žarković, S. S.; *Greške popisa stanovništva* [Population census errors], *Studije i analize* No. 3, Federal Statistical Office, Beograd 1954.
[10] Žarković, S. S.; Potpunost popisivanja [Completeness of enumeration], *Studije i analize* No. 4, Federal Statistical Office, Beograd 1954.
[11] Žarković, S. S.; Rezultati uzorka [Estimating census figures], *Studije i analize* No. 1, Federal Statistical Office, Beograd 1953.
[12] Žarković, S. S.; Sampling control of literacy data, *Journal of the American Statistical Association*, 49 (1954), pp. 510–519.
[13] Žarković, S. S., and Balaban, V.; Ispitivanje pismenosti [An inquiry into the literacy data], *Studije i analize* No. 7, Federal Statistical Office, Beograd 1955.

ON ADJUSTING SAMPLE TABULATIONS TO CENSUS COUNTS

M A EL-BADRY
Cairo University

AND

F. F. STEPHAN
Princeton University

INTRODUCTION

THERE are many occasions in which statisticians have to consider how they can make the best use of data that are obtained by sampling, particularly when they have available to them data from other sources that can be used to improve the estimates they wish to derive from the sample Problems of this kind are likely to require the attention of statisticians to an increasing extent as the flow of information from sample surveys and sample tabulations drawn from established record systems expands There are many indications of a strong trend in this direction It promises to provide more data, more quickly, and estimates of greater value than the older sources of a more cumbersome or more casual nature At the same time its value will depend on the fullest use of the data that are provided by other sources for the improvement of sample estimates and also for the supplementing of each by the other.

An outstanding example of this problem is presented by the numerous tabulations of data from samples that are part of the 1950 Census of Population In these the connection between the sample and the other source of data is very close since the sample is taken in connection with the regular enumeration or by a selection from the returns after they are completed. By these procedures the Bureau of the Census has greatly enriched the information it has been able to publish about the American population The introduction of an element of sampling error into the results is much more than offset by the value of the additional variables and cross tabulations that sampling made possible

While the effect of sampling error may be a very reasonable price to pay for the additional data that sampling provides, a statistician very naturally tries to reduce the cost to a minimum Moreover, he often finds it awkward to have two figures for the same item of information, one from the regular census count and the other from the sample estimate. He finds the public rather impatient in insisting that such figures be reconciled and that only the "correct" figure should be published.

For his own purposes, however, he is likely to find that he is better off with the original figures for the sample as well as the full count, since he can then make the best use of both for the particular problem on which he is working Just as he often finds after rounding a set of percentages that he must adjust them further to meet the public expectation that they total to exactly 100 per cent, so too, he often finds that he must make a somewhat artificial adjustment of sample estimates to make them add up to certain established totals He very naturally wishes to hold to a minimum the cost of such arbitrary adjustments and at the same time to reduce the effect of sampling error. This leads to a search for economical and effective methods of adjusting estimates from samples to make them agree in certain ways with other data and to reduce as far as is practicable the elements of sampling error that remain in the estimates after the adjustment.

The search for good methods is not simple The statistician recognizes that there are other sources of error than sampling in his data, that modern sampling is done quite often according to very complex designs, and that even with electronic equipment computing adjustments may be fairly expensive unless the procedures are simple. It may be worth while to sacrifice some part of the effectiveness of the adjustment in order to reduce the work to simple and economical methods that will be widely understood, consistent from one set of data to another, and readily reduced to routine At the same time it is not worth while to make needless sacrifices in the effectiveness of the methods and to this end it is possible to get some help from the application of statistical theory even when the full complexity of the problem is not adequately reflected in the theoretical models that are used. This paper is directed toward the development of the theory of some simple adjustment procedures.

AN EXAMPLE OF THE PROBLEM OF ADJUSTMENT

Some of the questions that arise in the adjustment of sample estimates may be illustrated by the figures in Table I for the Standard Metropolitan Area of Austin, Texas, showing school enrolment by age. For the population 5 and 6 years of age the sample agrees so closely with the census count that hardly any adjustment is needed to make them agree exactly In contrast, the sample estimate of the population 7 to 13 years old exceeds the count by 6 per cent and the sample estimate of the number of children enrolled in school is greater than the census count of all children of that age! This clearly shows the need for an adjustment of the estimates. It seems reasonable to make the ad-

justment by simply multiplying the count of the population of this age by the proportion of the children in the sample in this age group who were enrolled in school. This would correct most of the sampling error in the estimate of school enrolment. Similar but more moderate adjustments are needed for the ages 14–15 and 16–17. They can also be made for the older age groups. In the case of the older ages, however, the statistician may have noticed that there is evidence of a persistent tendency among the sample estimates for various areas to underestimate the number of adult males. This is the result of causes that are not fully understood Among the causes there appears to be a slight tendency among enumerators to avoid putting the head of a household on the sample lines predesignated on the census schedules. It may be associated with the fact that somewhat more information is required in the case of heads of households than in the case of other members of their families However, the listing procedures do not offer the enumerator any choice in the order of listing the head of the family and the tendency is probably limited to a minority of the enumerators

If this tendency is the source of the underestimate of the population 18 years old and older then the figures for school enrolment in these

TABLE I*

ADJUSTMENT OF ESTIMATES OF SCHOOL ENROLMENT, BY AGE, IN THE STANDARD METROPOLITAN AREA OF AUSTIN, TEXAS, IN 1950

Age	Total Population		Enrolled in School		
	Census count	Sample estimate	Sample estimate	Per cent	Adjusted estimate
5– 6	5,638	5,635	1,390	24.7	1,391
7–13	15,298	16,180	15,350	94 9	14,513
14–15	4,008	3,900	3,600	92 3	3,700
16–17	4,154	4,305	3,050	70.8	2,943
18–19	7,517	7,155	4,065	56 8	4,271
20–24	19,641	19,310	7,490	38 8	7,618
25–29	16,298	15,715	3,290	20.9	3,412

* The source of all the figures except the last column is *1950 Census of Population*, volume II, part 43, *Texas*, 43–81 and 43–98 The sample was taken by designating every fifth person listed by each enumerator to be included in the sample The sample estimates were formed by multiplying the sample tabulations by 5. For additional details about the sampling see the introductory text of any of the parts of the above reference. The adjusted estimate is formed by multiplying the sample estimate of children enrolled in school by the census count for the same age group and then dividing the product by the sample estimate of the population of that age group.

ages may be about right or even possibly overestimates. The effect of the adjustment in this case could be to increase the error of the estimates of enrolment. The problem is complicated by the high proportion of university students in this area and the likelihood that the effect of sampling was different to some degree for them than for school children almost all of whom were living at home. This example leaves unanswered, then, some of the questions that arise in the use of sample estimates in connection with other sources of data.

The method of adjustment that was used in Table I is the method that is suggested in the Census reports under the title "ratio estimates." It brings the estimates into agreement with one set of census counts but frequently there are several counts to which the sample estimates can be adjusted and the adjustment to one of them leaves the estimates out of agreement with the others. If it is important that the estimates be completely consistent with several sets of counts, the adjustment must be done simultaneously for all and the procedure becomes more complicated. In this case more help is needed from statistical theory to develop methods that tend to reduce the effects of sampling error at the same time that they succeed in adjusting the estimates so that certain totals of them agree with data from other sources

Before considering the application of theory, we may find it useful to examine the possibilities of extending the application of the ratio estimating procedure that was used in Table I to the adjustment of estimates to two sets of counts.

AN EXAMPLE OF TWO-WAY ADJUSTMENT

A good many of the Census sample tabulations consist of a cross classification of two variables from the sample for each of which the simple frequency distributions are available from the complete count. For example, the reports give for each state, standard metropolitan area, and city of 100,000 or more inhabitants a table derived from the 20 per cent sample showing the age distribution of each of the following groups:

(1) males in the labor force

(2) males in the civilian labor force who are employed

(3) males in the civilian labor force who are unemployed

(4) males who are not in the labor force and the same groups of females. The reports also show the complete counts of males and females in these groups and also the age distribution of all males and of all females. Children under 14 are excluded from all these tabula-

TABLE II

EMPLOYMENT STATUS ACCORDING TO AGE OF RURAL NON-FARM MALES IN A 20 PER CENT SAMPLE FOR NEW JERSEY AND THE CORRESPONDING CENSUS COUNTS BY AGE AND BY EMPLOYMENT STATUS FROM THE 1950 CENSUS OF POPULATION

Age	Military Labor Force (1)	Civilian Labor Force: Employed (2)	Civilian Labor Force: Unemployed (3)	Not in the Labor Force (4)	Total from Sample (5)	Total from Census Count (6)
14–19	7,740	4,760	730	14,335	27,565	27,063
20–24	6,660	11,705	990	3,220	22,575	22,822
25–29	3,215	15,445	680	2,575	21,915	22,761
30–34	2,170	17,160	505	1,770	21,605	22,555
35–39	1,095	16,925	485	1,730	20,235	21,285
40–44	385	15,670	460	1,665	18,180	19,388
45–49	170	13,600	440	1,830	16,040	16,736
50–54	95	11,940	570	2,530	15,135	15,761
55–59	30	9,670	615	2,960	13,275	13,845
60–64	10	6,985	485	3,175	10,655	11,621
65–69	0	4,155	415	4,175	8,745	9,087
70–74	10	2,055	145	4,340	6,550	6,700
75+	15	985	65	6,160	7,225	7,646
Total from sample	21,595	131,055	6,585	50,465	209,700	
Total from census count	21,830	137,075	6,532	51,833		217,270

tions The nature of the adjustments that can be made in the sample tabulations from the census counts can be seen in the examples given in Tables II, III, and IV

In Table II the right hand column and bottom row show the counts to which the estimates are to be adjusted It is evident, especially in the case of the estimates of the number of unemployed that adjustment by rows will give one quite different estimates than the adjustment by columns if the ratio method that was used in Table I is applied to this table The figures for the number of unemployed would differ by about 3½ per cent The other estimates would be much closer to agreement If the adjustment is made by rows and then another adjustment by columns is applied to the resulting table of estimates, the new row totals will no longer agree with the corresponding census counts. From this it is clear that the ratio method cannot be used to bring all the totals into agreement with the census counts simultaneously, except by some modification of the procedure

Table III shows the results of applying the ratio adjustment by rows The column totals can now be brought into agreement with the cor-

TABLE III

RESULTS OF MAKING RATIO ESTIMATE ADJUSTMENTS OF THE ROWS IN TABLE II

Age	Ratio for Adjustment of Row	Estimates after Adjustment by Rows				Adjusted Row Totals	Proportion for Second Adjustment
		(1)	(2)	(3)	(4)		
14–	.981788	7,599.0	4,673.3	716.7	14,073.9	27,063	.124559
20–	1.010941	6,732.9	11,833.1	1,000.8	3,255.2	22,822	.105040
25–	1.038604	3,339.1	16,041.2	706.3	2,674.4	22,761	.104759
30–	1.043971	2,265.4	17,914.5	527.2	1,847.8	22,555	.103811
35–	1.051890	1,151.8	17,803.2	510.2	1,819.8	21,285	.097966
40–	1.066447	410.6	16,711.2	490.6	1,775.6	19,388	.089235
45–	1.043392	177.4	14,190.1	459.1	1,909.4	16,736	.077029
50–	1.041361	98.9	12,433.9	593.6	2,634.6	15,761	.072541
55–	1.042938	31.3	10,085.2	641.4	3,087.1	13,845	.063723
60–	1.090662	10.9	7,618.3	529.0	3,462.9	11,621	.053486
65–	1.039108	0	4,317.5	431.2	4,338.3	9,087	.041824
70–	1.022901	10.2	2,102.1	148.3	4,439.4	6,700	.030837
75+	1.058270	15.9	1,042.4	68.8	6,518.9	7,646	.035191
Total		21,843.4	136,766.0	6,823.2	51,837.3	217,270	
Count		21,830	137,075	6,532	51,833	217,270	
Difference		13.4	−309.0	291.2	4.3		

responding counts by modifying the adjustment in a manner that will maintain the adjustment by rows. This can be done by distributing the needed total adjustment in each column in proportion not to the cell frequencies but to the adjusted row totals. The results are shown in Table IV, as the upper of the two figures in each cell.

If the adjustment had been made first by columns and then using the modification of the ratio method by rows, the results would be those shown by the lower of the two figures in each cell of Table IV. A comparison of the two estimates in each cell shows the degree to which the two ways of applying the adjustments differ. The average of the two estimates will tend to be better than either of them alone.

The modification of the ratio method that was used in this example is a particular case of a device given in [5, p. 169, step (11)] and applied to the first adjustment as well as the second in [4, p. 249]. The effectiveness of this method of adjustment will be examined in a later section of this article.

The method of making the adjustment may be expressed in general terms as follows:

(1) Let the sample frequency in the ith row and jth column of the table be n_{ij} and the totals of the ith row and jth column be $n_{i.}$ and $n_{.j}$

respectively. Let $N_{i.}$ and $N_{.j}$ be the row and column totals from the census count. Finally let n and N be the grand totals of the sample and the census count respectively. Then the figures in Table II are formed by multiplying each n_{ij} by N/n. In Table III the ratios for the adjustment of rows are computed by taking $N_{i.}$ and dividing it by the ith row total which is equal to $n_i\,(N/n)$.

(2) The cell frequencies of the adjusted table such as those shown in columns (1) to (4) of Table III are then computed by multiplying the estimates in Table II by the ratios for their respective rows. (These results can be obtained directly by multiplying the original sample frequencies n_{ij} by $N_{i.}/n_{i.}$.) The right hand column of Table III can be computed from the census counts or from the row totals since they now agree. The column totals are obtained and the differences found to complete Table III.

(3) For each column the difference between its total and the corresponding census count is multiplied by each of the proportions for the second adjustment in turn and the products are then subtracted from the numbers on the corresponding row in the same column. This yields the final estimates which are equal to

$$n_{ij}(N_{i.}/n_{i.}) - (N_{i.}/N)\left[\sum (n_{hj}N_{h.}/n_{h.}) - N_{.j}\right] \qquad (1)$$

the summation being over $h=1, 2, \cdots, r$ for a table of r rows. A similar method is used with rows and columns interchanged when the first adjustment is made by columns and the second by rows.

APPLICATION OF STATISTICAL THEORY TO THE ADJUSTMENT OF SAMPLE ESTIMATES

Statistical theory offers several approaches to the problem of finding suitable methods of adjusting sample estimates. Among the principal approaches are those of the classical least squares theory as developed by Markoff, the principle of maximum likelihood developed by Fisher, and the approach of decision theory developed by Wald. The application of any of these approaches is complicated by the fact that the sampling is usually done with a considerable number of departures from simple random sampling, such as stratification, systematic selection, the use of clusters as units of sampling, subsampling, variable probabilities of selection, weighting, etc. Consequently a model that faithfully represents the sampling process may be too complex to be utilized in any of these approaches. Moreover, the actual sampling operations may be affected by biases and eccentricities other than those that are inherent in the model. For these reasons a moderately simple

TABLE IV

FINAL ADJUSTMENT OF THE ESTIMATES IN TABLE II DISTRIBUTING COLUMN DIFFERENCES IN PROPORTION TO ROW TOTALS (UPPER FIGURES) AND CORRESPONDING ADJUSTMENT MADE INTERCHANGING COLUMNS AND ROWS (LOWER FIGURES)

Age	Military Labor Force (1)	Civilian Labor Force: Employed (2)	Civilian Labor Force: Unemployed (3)	Not in the Labor Force (4)	Total
14–	7,598	4,712	680	14,073	27,063
	7,705	4,229	688	14,441	
20–	6,732	11,865	970	3,255	22,822
	6,688	11,964	968	3,202	
25–	3,338	16,073	676	2,674	22,761
	3,254	16,178	675	2,654	
30–	2,264	17,947	497	1,847	22,555
	2,203	18,008	504	1,840	
35–	1,150	17,834	482	1,819	21,285
	1,129	17,840	488	1,828	
40–	409	16,739	465	1,775	19,388
	434	16,669	470	1,815	
45–	176	14,214	437	1,909	16,736
	174	14,239	437	1,886	
50–	98	12,456	572	2,635	15,761
	97	12,496	566	2,602	
55–	30	10,105	623	3,087	13,845
	35	10,146	612	3,052	
60–	10	7,635	513	3,463	11,621
	67	7,661	498	3,395	
65–	0	4,330	419	4,338	9,087
	4	4,372	413	4,298	
70–	10	2,112	139	4,439	6,700
	4	2,111	142	4,443	
75+	15	1,053	59	6,519	7,646
	36	1,162	71	6,377	
Total	21,830	137,075	6,532	51,833	217,270

and practical method of adjustment is attainable only after some considerable compromise with the known features of the sampling process that produced the data and also those of the methods of estimation which may have been employed One is forced to approximate the actual process by a greatly simplified model in order to find a feasible basis for applying statistical theory to the choice of a method of adjustment.

The application of statistical theory to the adjustment of sample estimates is further complicated by the fact that even when the errors of the sample data in their original form are independent, once the estimates are subjected to conditions that equate certain sums of estimates to given constants, the errors of the estimates are correlated. More than this, they form a linearly dependent set and it is necessary to either introduce additional variables in the form of Lagrange multipliers or reduce the set of estimation errors to a subset that is linearly independent before the application of the model proceeds.

It is clear that the degree of success that has been attained by previous efforts such as those of Deming and Stephan [1] in utilizing least squares and by Smith [4] in applying maximum likelihood depended on the use of a relatively simple model to represent the actual sampling process, namely the *multinomial* model. This model assumes sampling with replacement (or from an infinite universe) and is characterized by the *parameters* p_{ij} which are the probabilities that an element (person) drawn for the sample will be of the kind that is classified on the ith row and in jth column. The magnitudes of the parameters are not known but are to be estimated from the data they produce Stephan's iterative procedure [5] which neglected the correlations among the errors can be used to apply the approach of simple decision theory when the loss function for the adjustment can be approximated well enough by a linear function of the squares of the amounts by which the first estimates of population cell frequencies, $(N/n)n_{ij}$, are adjusted This may be a substantial compromise since it attains a practical method at the cost of possibly inaccurate formulation of the loss function Further inquiries are needed to develop simple methods with a better theoretical foundation that take account of the relevant properties of the actual sampling process

Even when the problem is simplified by the assumption that the sampling conforms to the multinomial model, the actual computing of the adjustments may be laborious and other difficulties may be encountered. It may be instructive to examine the successive steps in

the application of the least squares approach, starting with Markoff's theorem.

MARKOFF ESTIMATES

Markoff's theorem on estimation deals with problems of point estimation in which a number of observations, here represented by $\|X\|$, i e , the vector $[X_1, X_2, \cdots, X_n]$, are known to have expected values that are linear functions of the parameters to be estimated, $\|t\|$, and it is desired to obtain estimates which are linear functions of the observations, $\|X\|$.

The theorem states that if the n readings $\|X\|$ are uncorrelated and have the same variance and if the relations between the expected values of the X's and the m unknown parameters $\|t\|$ $(m \leqq n)$ can be expressed in the form of a set of n linear equations of design

$$E\|X\| = \|A\| \, \|t\| \tag{2}$$

where $\|A\|$ is an n by m matrix of known elements, then the estimates $\|T\|$ of $\|t\|$ which are 1) linear in the observations, 2) unbiased, and 3) of minimum variance subject to 1 and 2 are given by

$$\|T\| = \| \, \|A\|'\|A\| \, \|^{-1} \|A\|'\|X\| \tag{3}$$

providing $\|A\|' \, \|A\|$ is nonsingular

In the case in which the readings are correlated the theorem takes the form [6]

$$\|T\| = \| \, \|A\|'\|\sigma^{gh}\| \, \|A\| \, \|^{-1}\|A\|'\|\sigma^{gh}\| \, \|X\| \tag{4}$$

where $\|\sigma^{gh}\|$ is the inverse of the matrix of covariances $\|\sigma_{gh}\|$, provided the covariance matrix is nonsingular.

It may happen that the variances and covariances σ_{gh} are functions of the unknown parameters $\|t\|$ and that consequently the equations of estimation (4) do not give explicit formulas for the estimates $\|T\|$. In such cases it is sometimes possible to obtain practical approximations to the Markoff estimates [6] by substituting for the needed numerical values of $\|t\|$ in the functions σ^{gh} the symbols for the estimates of these values $\|T\|$ and solving the resulting equations for $\|T\|$. The equations to be solved will usually be quite difficult and some approximate method for their solution, commonly iterated to improve the approximation, will be used These estimates give the values of the parameters for which, *given the sample*, a certain function of the parameters is minimized. In this it resembles the principle of maximum likelihood

Indeed, for those probability models in which minimizing this function also maximizes the likelihood, the estimates are maximum likelihood estimates. These practical approximations are, then,

$$\|\hat{T}\| = \| \|A\|'\|\hat{\sigma}^{gh}\| \|A\| \|^{-1}\|A\|'\|\hat{\sigma}^{gh}\| \|X\| \tag{5}$$

where $\hat{\sigma}^{gh}$ has the same algebraic form as σ^{gh} with $\|t\|$ replaced by $\|\hat{T}\|$.

Let us consider now in general terms the application of Markoff's theorem to the case of contingency table of r rows and s columns. Suppose that the table is the result of taking a random sample of n readings in conformity with a multinomial model in which the probability corresponding to the cell of the table on the ith row and in the jth column is p_{ij} The number of readings n_{ij} that are observed to fall into the ijth cell is multiplied by N/n in order to obtain estimates X_{ij} of the population frequencies in a population of size N. We will discuss first the case in which none of the marginal totals of the population are known, i e , only N is known. Now if we express the expected values N_{ij} of the X_{ij} as linear functions of $rs-1$ unknown parameters t_i in the form

$$N_{ij} = a_{ij}t_1 + b_{ij}t_2 + c_{ij}t_3 + \cdots + d_{ij} \tag{6}$$

where a_{ij}, b_{ij}, $\cdots$, are known numbers given by the conditions of the problem and set forth to estimate these parameters we will notice that, before applying Markoff's theorem, the last equation giving the expected value of X_{rs} in terms of the t's must be discarded because it is merely a linear combination of the preceding $rs-1$ equations Also, since in the case of a multinomial distribution[1] var $X_{ij} \propto p_{ij}(1-p_{ij})$ (where $p_{ij} = N_{ij}/N$) and cov $(X_{ij}, X_{kl}) \propto -p_{ij}p_{kl}$, it can easily be shown that the elements of $\|\sigma^{gh}\|$ are:

$$\sigma^{gg} \propto \frac{1}{N_{rs}} + \frac{1}{N_{gg}}, \qquad \sigma^{gh} \propto \frac{1}{N_{rs}} \quad (p \neq q),$$

$p, q = 1, 2, \cdots, rs-1$.

Now since

$$\sum_{i=1}^{r}\sum_{j=1}^{s} N_{ij} = N,$$

whatever values the t may take, we must have

$$\sum_i \sum_j a_{ij} = \sum_i \sum_j b_{ij} = \cdots = 0$$

[1] It is assumed here that if the population is finite, its size N is large enough to permit the use of the multinomial model.

and

$$\sum_i \sum_j d_{ij} = N.$$

This means that in the first $rs-1$ equations of design $\sum a_{ij} = -a_{rs}$, $\sum b_{ij} = -b_{rs}, \cdots$, and $\sum d_{ij} = N - d_{rs}$. We also have $\sum X_{ij} = N - X_{rs}$. Using these relations we see that $\|A\|' \, \|\sigma^{gh}\| \, \|A\|$ and $\|A\|' \, \|\sigma^{gh}\| \, \|X\|$ finally take the forms

$$\|A\|'\|\sigma^{gh}\| \, \|A\| \propto \left\| \begin{matrix} \sum \dfrac{a_{ij}^2}{N_{ij}}, & \sum \dfrac{a_{ij}b_{ij}}{N_{ij}}, & \sum \dfrac{a_{ij}c_{ij}}{N_{ij}}, & \cdots \\ \sum \dfrac{b_{ij}a_{ij}}{N_{ij}}, & \sum \dfrac{b_{ij}^2}{N_{ij}}, & \sum \dfrac{b_{ij}c_{ij}}{N_{ij}}, & \cdots \\ \sum \dfrac{c_{ij}a_{ij}}{N_{ij}}, & \sum \dfrac{c_{ij}b_{ij}}{N_{ij}}, & \sum \dfrac{c_{ij}^2}{N_{ij}}, & \cdots \\ \vdots & \vdots & \vdots & \end{matrix} \right\|$$

and

$$\|A\|'\|\sigma^{gh}\| \, \|X\| \propto \left\| \begin{matrix} \sum \dfrac{a_{ij}(X_{ij} - d_{ij})}{N_{ij}} \\ \sum \dfrac{b_{ij}(X_{ij} - d_{ij})}{N_{ij}} \\ \sum \dfrac{c_{ij}(X_{ij} - d_{ij})}{N_{ij}} \\ \vdots \end{matrix} \right\|$$

the summations being over $i = 1, 2, \cdots, r; j = 1, 2, \cdots, s$.

The equations (4) can then be written in the form:

$$\begin{aligned} T_1 \sum \frac{a_{ij}^2}{N_{ij}} + T_2 \sum \frac{a_{ij}b_{ij}}{N_{ij}} + \cdots &= \sum \frac{a_{ij}(X_{ij} - d_{ij})}{N_{ij}}, \qquad (7) \\ T_1 \sum \frac{b_{ij}a_{ij}}{N_{ij}} + T_2 \sum \frac{b_{ij}^2}{N_{ij}} + \cdots &= \sum \frac{b_{ij}(X_{ij} - d_{ij})}{N_{ij}} \\ &\vdots \end{aligned}$$

$(rs - 1$ equations)

or

$$\sum \frac{a_{ij}}{N_{ij}}(X_{ij} - N_{ij}') = 0, \quad \text{where}$$

$$N_{ij}' = a_{ij}T_1 + b_{ij}T_2 + c_{ij}T_3 + \cdots + d_{ij},\cdot$$

$$\sum \frac{b_{ij}}{N_{ij}}(X_{ij} - N_{ij}') = 0$$

$(rs - 1$ equations).

The practical approximations referred to in equation (5) will be given by the set of $rs-1$ simultaneous equations:

$$\sum \frac{a_{ij}X_{ij}}{\widehat{N}_{ij}} = 0, \text{ where } \widehat{N}_{ij} = a_{ij}\widehat{T}_1 + b_{ij}\widehat{T}_2 + c_{ij}\widehat{T}_3 + \cdots + d_{ij}$$

$$\sum \frac{b_{ij}X_{ij}}{\widehat{N}_{ij}} = 0 \tag{8}$$

$$\vdots$$

which are evidently equivalent to the maximum likelihood equations of estimation The simultaneous solution of equations (8) will give the approximate Markoff estimates $\|\widehat{T}\|$. Up to this point we have been considering the case in which only N is known When the column totals of the expected values are known, the above setup remains unchanged except that we must have $\sum_i a_{ij} = \sum_i b_{ij} = \cdots = 0$, and $\sum_i d_{ij} = N_{\cdot j}$. Also the number of parameters under estimation is reduced to $(r-1)s$ When the row totals are also known we must have in addition to the above constraints $\sum_j a_{ij} = \sum_j b_{ij} = \cdots = 0$ and $\sum_j d_{ij} = N_{i\cdot}$ and the number of parameters is reduced to $(r-1)(s-1)$.

ESTIMATION OF POPULATION CELL FREQUENCIES

a) *Two-way table with one set of population marginal totals known* If the column totals are known then we take the unknown cell frequencies N_{ij} in the upper $r-1$ rows of the table as our $(r-1)\times s$ unknown parameters The observations are $X_{ij} = (N/n)n_{ij}$ of which the expected values are N_{ij}. Equations (8) become

$$\sum \frac{a_{ij}n_{ij}}{\widehat{N}_{ij}} = 0,\ \sum \frac{b_{ij}n_{ij}}{\widehat{N}_{ij}} = 0, \cdots (r-1) \times s \text{ equations} \tag{9}$$

where

$$a_{11} = 1, \quad a_{r1} = -1, \quad a_{i1} = 0 \ (i \neq 1 \text{ or } r),$$
$$b_{12} = 1, \quad b_{r2} = -1, \quad a_{i2} = 0 \ (i \neq 2 \text{ or } r),$$

i.e.,

$$\frac{\widehat{N}_{11}}{n_{11}} = \frac{\widehat{N}_{21}}{n_{21}} = \frac{\widehat{N}_{31}}{n_{31}} = \cdots = \frac{\widehat{N}_{r1}}{n_{r1}} = \frac{\widehat{N}_{.1}}{n_{.1}},$$
$$\frac{\widehat{N}_{i2}}{n_{i2}} = \frac{\widehat{N}_{.2}}{n_{.2}} \quad (i = 1, 2, \cdots, r).$$

This shows that the estimates of the population cell frequencies in any column are proportional to the observations in the same column.

b) *Two-way table with the two sets of population marginal totals known.* In this case we take as our unknown parameters the population frequencies in the first $r-1$ rows and $s-1$ columns Equations (8) then become

$$\sum \frac{a_{ij}n_{ij}}{\widehat{N}_{ij}} = 0, \quad \sum \frac{b_{ij}n_{ij}}{\widehat{N}_{ij}} = 0, \cdots \quad ((r-1)(s-1) \text{ equations})$$

where

$a_{11}=1$, $a_{1s}=-1$, $a_{r1}=-1$, $a_{rs}=1$, and all the remaining a_{ij}'s vanish,

$b_{12}=1$, $b_{1s}=-1$, $b_{r2}=-1$, $b_{rs}=1$, and all the remaining b_{ij}'s vanish,

$c_{13}=1$, $c_{1s}=-1$, $c_{r3}=-1$, $c_{rs}=1$, and all the remaining c_{ij}'s vanish.

The approximate Markoff estimates are therefore given by the $(r-1)$ $(s-1)$ equations

$$\begin{aligned} \frac{n_{11}}{\widehat{N}_{11}} - \frac{n_{1s}}{\widehat{N}_{1s}} - \frac{n_{r1}}{\widehat{N}_{r1}} + \frac{n_{rs}}{\widehat{N}_{rs}} &= 0, \\ \frac{n_{12}}{\widehat{N}_{12}} - \frac{n_{1s}}{\widehat{N}_{1s}} - \frac{n_{r2}}{\widehat{N}_{r2}} + \frac{n_{rs}}{\widehat{N}_{rs}} &= 0, \end{aligned} \tag{10}$$

THE METHOD OF PROPORTIONAL DISTRIBUTION OF MARGINAL ADJUSTMENTS

In the foregoing applications of least squares or maximum likelihood the statistician is frequently presented with a considerable number of equations to be solved simultaneously. It is natural under these conditions to look for short cuts and approximate methods. One such that has been suggested for use when the cell frequencies are roughly proportional is the method of proportional distribution of marginal adjustments [4]. It proceeds to compute the adjusted estimates M_{ij} of the population frequency in the ijth cell,

$$M_{ij} = \frac{N}{n}\left\{n_{ij} + \frac{N_{.j}}{N}\left(\frac{n}{N} N_{i.} - n_{i.}\right) + \frac{N_{i.}}{N}\left(\frac{n}{N} N_{.j} - n_{.j}\right)\right\}. \quad (11)$$

This estimate is unbiased and adds up to the known marginal totals of the population. M_{ij} has also the advantage of simplicity of processing. Yet, in the case of small cell probabilities, the variance of M_{ij} may be larger than that of the original inflated estimate $I_{ij} = (N/n)n_{ij}$. To illustrate this result consider the variance of M_{ij}. Since cov $(n_{ij}, n_{i.}) = np_{ij}q_{i.}$, cov $(n_{ij}, n_{.j}) = np_{ij}q_{.j}$, and cov $(n_{i.}, n_{.j}) = -np_{i.}p_{.j} + np_{ij}$, the variance of M_{ij} is

$$\text{var } M_{ij} = \frac{N^2}{n} p_{ij}q_{ij} + \frac{N^2}{n}\left\{2p_{ij}(3p_{i.}\, p_{.j} - p_{i.} - p_{.j}) + p_{i.}\, p_{.j}(p_{i.} + p_{.j} - 4p_{i.}\, p_{.j})\right\}.$$

The first term on the right hand side of this equation is the variance of I_{ij}. Therefore the variance of M_{ij} will be greater than that of I_{ij} when

$$\{4p_{i.}\, p_{.j} - (p_{i.} + p_{.j})\}(2p_{ij} - p_{i.}\, p_{.j}) > 2p_{ij}p_{i.}p_{.j}$$

i.e.,

$$\{4 - u_{ij}\}\left(1 - \tfrac{1}{2}p_{i.}\, p_{.j}/p_{ij}\right) > 1 \quad (12)$$

where

$$u_{ij} = \frac{1}{p_{i.}} + \frac{1}{p_{.j}}.$$

Now u_{ij} can take any value between 2 and infinity but the above inequality (12) is not satisfied by the values of u_{ij} ranging between 3 and 4 because then $[1 - (p_{i.}\, p_{.j}/2p_{ij})]$ would have to be greater than a quantity which is itself greater than or equal to one. This is naturally impossible because $p_{i.}\, p_{.j}/2p_{ij}$ is always positive. The inequality is not

satisfied for values of u_{ij} ranging between 2 and 3 either because in that case it would be satisfied only if

$$\frac{2p_{ij}}{p_{i.}p_{.j}} - 1 > \frac{1}{3 - u_{ij}}. \tag{13}$$

Assuming that $p_{.j}$ is the larger of $p_{i.}$ and $p_{.j}$ and since $p_{ij} \leqq p_{i.}$ we must have

$$\frac{2}{p_{.j}} - 1 > \frac{1}{3 - u_{ij}}.$$

But evidently

$$\frac{2}{p_{.j}} - 1 < \frac{1}{3 - \dfrac{2}{p_{.j}}} < \frac{1}{3 - u_{ij}}\,.$$

and consequently (13) cannot be satisfied.

Thus the variance of M_{ij} is larger than that of I_{ij} when

$$4 < u_{ij} < \infty \quad \text{and} \quad p_{ij} < \frac{p_{i.}p_{.j}}{2}\left(1 - \frac{1}{u_{ij} - 3}\right). \tag{14}$$

Equation (14) is likely to take place in tables in which some of the marginal totals are small

For instance in a population of 100,000 individuals if the row and column totals corresponding to a certain cell are 40,000 each then M_{ij} would have a larger variance than I_{ij} if the population frequency in the cell is less than 4,000 If the marginal frequencies are 10,000 each then var $M_{ij}>$var I_{ij} when the population frequency in the cell is 470 or less. It should also be mentioned that this drawback is independent of the sample size and hence would not be cured by increasing n.

THE METHOD OF ARRAY PROPORTIONS

In this section estimates based on $n_{ij}/n_{i.}$ and $n_{ij}/n_{.j}$, the row and column proportions, will be considered. Before discussing the estimates we have to study certain properties of these proportions We are going to discard the case where the array total in the sample is zero. In practice, that would mean that if the sample table has a certain array total of zero, we would not proceed to estimate the population cell frequencies in the corresponding population array.

The expected value of an array proportion.

$$E\left(\frac{n_{ij}}{n_{i.}}(n_{i.}\neq 0)\right)=\sum_{n_{i.}=1}^{n}\left[Pr\{n_{i.}(n_{i.}\neq 0)\}\,E\left(\frac{n_{ij}}{n_{i.}}\mid n_{i.}(n_{i.}\neq 0)\right)\right]$$

$$=\sum_{n_{i.}=1}^{n}Pr\{n_{i.}(n_{i.}\neq 0)\}\,\frac{p_{ij}}{p_{i.}}(1-q_{i.}{}^{n})^{-1} \tag{15}$$

$$=\frac{p_{ij}}{p_{i.}}(1-q_{i.}{}^{n})^{-1}(1-q_{i.}{}^{n})=\frac{p_{ij}}{p_{i.}}.$$

Row (or column) proportions not belonging to the same row (or column) are uncorrelated.

To prove this result it will be shown that

$$E\frac{n_{ij}}{n_{i.}}\,\frac{n_{kl}}{n_{k.}}=E\frac{n_{ij}}{n_{i.}}\,E\frac{n_{kl}}{n_{k.}},$$

where $k\neq i$ and $n_{i.}, n_{k.}\neq 0$

$$E\left(\frac{n_{ij}}{n_{i.}}\,\frac{n_{kl}}{n_{k.}}\right)=\sum_{n_{i.},n_{k.}=1}^{n}\left[Pr\{n_{i.}, n_{k.}\}E\left(\frac{n_{ij}}{n_{i.}}\,\frac{n_{kl}}{n_{k.}}\mid n_{i.}, n_{k.}\right)\right].$$

on both sides of which it is understood that $n_{i.}, n_{k.}\neq 0$.

Hence

$$E\left(\frac{n_{ij}}{n_{i.}}\,\frac{n_{kl}}{n_{k.}}\right)=\sum_{n_{i.},n_{k.}}^{n}Pr\{n_{i.}, n_{k.}\}(1-Pr\{n_{i.}, n_{k.}\neq 0\})^{-1}\frac{p_{ij}}{p_{i.}}\,\frac{p_{kl}}{p_{k.}}$$

$$=(1-Pr\{n_{i.}, n_{k.}\neq 0\})^{-1}\frac{p_{ij}}{p_{i.}}\,\frac{p_{kl}}{p_{k.}}(1-Pr\{n_{i.},n_{k.}\neq 0\})$$

$$=\frac{p_{ij}}{p_{i.}}\,\frac{p_{kl}}{p_{k.}}=E\frac{n_{ij}}{n_{i.}}\,E\frac{n_{kl}}{n_{k.}}.$$

Evidently this relation can be extended to any number of proportions

We thus see that

$$\operatorname{cov}(n_{ij}/n_{i.}, n_{kl}/n_{k.})=0. \tag{16}$$

Correlation of array proportions belonging to the same row (or column)

Adopting the multinomial model, i.e., assuming that the population is large, we have

$$Pr\{n_{ij}, n_{il}, n_{i.}\}=Pr\{n_{i.}\}Pr\{n_{ij}, n_{il}\mid n_{i.}\}$$

$$= \frac{n!(p_{i\cdot})^{n_{i\cdot}}(1-p_{i\cdot})^{n-n_{i\cdot}}}{(n_{i\cdot})!(n-n_{i\cdot})!} \; \frac{(n_{i\cdot})!\left(\dfrac{p_{ij}}{p_{i\cdot}}\right)^{n_{ij}}\left(\dfrac{p_{il}}{p_{i\cdot}}\right)^{n_{il}}}{(n_{ij})!(n_{il})!(n_{i\cdot}-n_{ij}-n_{il})!}$$
$$\cdot\left(1-\frac{p_{ij}+p_{il}}{p_{i\cdot}}\right)^{n_{i\cdot}-n_{ij}-n_{il}}.$$

Hence

$$E\,\frac{n_{ij}}{n_{i\cdot}}\,\frac{n_{il}}{n_{i\cdot}} = \frac{p_{ij}}{p_{i\cdot}}\,\frac{p_{il}}{p_{i\cdot}}(1-q_{i\cdot}{}^{n})^{-1}\sum_{n_{i\cdot}=1}^{n}\frac{n_{i\cdot}-1}{n_{i\cdot}}\,\frac{n!(p_{i\cdot})^{n_{i\cdot}}(1-p_{i\cdot})^{n-n_{i\cdot}}}{(n_{i\cdot})!(n-n_{i\cdot})!}$$
$$\sum_{n_{ij},n_{il}=1}^{n_{i\cdot}-1}\frac{(n_{i\cdot}-2)!\left(\dfrac{p_{ij}}{p_{i\cdot}}\right)^{n_{ij}-1}\left(\dfrac{p_{il}}{p_{i\cdot}}\right)^{n_{il}-1}}{(n_{ij}-1)!(n_{il}-1)!(n_{i\cdot}-n_{ij}-n_{il})!}$$
$$\left(1-\frac{p_{ij}+p_{il}}{p_{i\cdot}}\right)^{n_{i\cdot}-n_{ij}-n_{il}}$$
$$= \frac{p_{ij}}{p_{i\cdot}}\,\frac{p_{il}}{p_{i\cdot}}(1-q_{i\cdot}{}^{n})^{-1}\left\{(1-q_{i\cdot}{}^{n})\right.$$
$$\left.-\sum_{n_{i\cdot}=1}^{n}\frac{n!(p_{i\cdot})^{n_{i\cdot}}(1-p_{i\cdot})^{n-n_{i\cdot}}}{n_{i\cdot}(n_{i\cdot})!(n-n_{i\cdot})!}\right\}$$
$$= \frac{p_{ij}}{p_{i\cdot}}\,\frac{p_{il}}{p_{i\cdot}}\left(1-E\frac{1}{n_{i\cdot}}\right) \qquad (n_{i\cdot}\neq 0)$$

and consequently

$$\operatorname{cov}\left(\frac{n_{ij}}{n_{i\cdot}},\frac{n_{il}}{n_{i\cdot}}\right) = -\frac{p_{ij}}{p_{i\cdot}}\,\frac{p_{il}}{p_{i\cdot}}\,E\,\frac{1}{n_{i\cdot}}. \tag{17}$$

Variance of the array proportion $n_{ij}/n_{i\cdot}$.
Since

$$Pr\{n_{ij},n_{i\cdot}\} = \frac{n!p_{i\cdot}{}^{n_{i\cdot}}(1-p_{i\cdot})^{n-n_{i\cdot}}}{(n_{i\cdot})!(n-n_{i\cdot})!}$$
$$\times\frac{(n_{i\cdot})!}{(n_{ij})!(n_{i\cdot}-n_{ij})!}\left(\frac{p_{ij}}{p_{i\cdot}}\right)^{n_{ij}}\left(1-\frac{p_{ij}}{p_{i\cdot}}\right)^{n_{i\cdot}-n_{ij}}$$

then

$$E\left(\frac{n_{ij}}{n_{i.}}\right)^2 = \frac{p_{ij}}{p_{i.}}(1-q_{i.}^n)^{-1}\sum_{n_{i.}=1}^{n}\frac{1}{n_{i.}}\frac{n!p_{i.}^{n_{i.}}(1-p_{i.})^{n-n_{i.}}}{(n_{i.})!(n-n_{i.})!}$$

$$\times\sum_{n_{ij}=1}^{n_{i.}} n_{ij}\frac{(n_{i.}-1)!\left(\frac{p_{ij}}{p_{i.}}\right)^{n_{ij}-1}}{(n_{ij}-1)!(n_{i.}-n_{ij})!}\left(1-\frac{p_{ij}}{p_{i.}}\right)^{n_{i.}-n_{ij}}.$$

The second summation

$$= 1+\sum_{n_{ij}=2}^{n_{i.}}\frac{(n_{i.}-1)!\left(\frac{p_{ij}}{p_{i.}}\right)^{n_{ij}-1}}{(n_{ij}-2)!(n_{i.}-n_{ij})!}\left(1-\frac{p_{ij}}{p_{i.}}\right)^{n_{i.}-n_{ij}},$$

$$= 1+\frac{p_{ij}}{p_{i.}}(n_{i.}-1)\sum_{n_{ij}=2}^{n_{i.}}\frac{(n_{i.}-2)!\left(\frac{p_{ij}}{p_{i.}}\right)^{n_{ij}-2}}{(n_{ij}-2)!(n_{i.}-n_{ij})!}\left(1-\frac{p_{ij}}{p_{i.}}\right)^{n_{i.}-n_{ij}}$$

$$= 1+(n_{i.}-1)\frac{p_{ij}}{p_{i.}} = n_{i.}\frac{p_{ij}}{p_{i.}}+1-\frac{p_{ij}}{p_{i.}},$$

and thus

$$E\left(\frac{n_{ij}}{n_{i.}}\right)^2 = \frac{p_{ij}}{p_{i.}}(1-q_{i.}^n)^{-1}\left\{\frac{p_{ij}}{p_{i.}}(1-q_{i.}^n)\right.$$

$$\left.+\left(1-\frac{p_{ij}}{p_{i.}}\right)(1-q_{i.}^n)E\frac{1}{n_{i.}}\right\}$$

$$= \left(\frac{p_{ij}}{p_{i.}}\right)^2+\frac{p_{ij}}{p_{i.}}\left(1-\frac{p_{ij}}{p_{i.}}\right)E\frac{1}{n_{i.}}.$$

Hence

$$\operatorname{var}\frac{n_{ij}}{n_{i.}} = \frac{p_{ij}}{p_{i.}}\left(1-\frac{p_{ij}}{p_{i.}}\right)E\frac{1}{n_{i.}}. \tag{18}$$

Estimates based on array proportions

Consider the estimates

$$L_{ij} = \frac{n_{ij}}{n_{.j}}N_{.j}+\frac{N_{.j}}{N}\left(N_{i.}-\sum_j\frac{n_{ij}}{n_{.j}}N_{.j}\right).^{2} \tag{19}$$

[2] In this case we could take as our estimate either L_{ij} or

$$L_{ij}' = \frac{n_{ij}}{n_{i.}}N_{i.}+\frac{N_{i.}}{N}\left(N_{.j}-\sum_i\frac{n_{ij}}{n_{i.}}N_{i.}\right)$$

The average of L_{ij} and L_{ij}' will have the advantage that its standard deviation is smaller than the average of those of L_{ij} and L_{ij}'.

The expected value of

$$L_{ij} = \frac{p_{ij}}{p_{.j}} N_{.j} + \frac{N_{.j}}{N}\left(N_{i.} - \sum_j \frac{p_{ij}}{p_{.j}} N_{.j}\right)$$

$$= N_{ij} + \frac{N_{.j}}{N}(N_{i.} - N_{i.}) = N_{ij}.$$

L_{ij} is therefore unbiased. It also adds up to the column and row totals since

$$L_{.j} = \sum_i L_{ij} = N_{.j} + N_{.j} - N_{.j} = N_{.j}$$

and

$$L_{i.} = \sum_j L_{ij} = \sum_j \frac{n_{ij}}{n_{.j}} N_{.j} + N_{i.} - \sum_j \frac{n_{ij}}{n_{.j}} N_{.j} = N_{i.}.$$

The sampling variance of L_{ij} is

$$\operatorname{var} L_{ij} = (N_{.j})^2 \operatorname{var} \frac{n_{ij}}{n_{.j}} + \left(\frac{N_{.j}}{N}\right)^2 \operatorname{var}\left(\sum_j \frac{n_{ij}}{n_{.j}} N_{.j}\right)$$

$$- 2\frac{(N_{.j})^2}{N} \operatorname{cov}\left(\frac{n_{ij}}{n_{.j}}, \sum_j \frac{n_{ij}}{n_{.j}} N_{.j}\right)$$

but

$$\operatorname{var}\left(\sum_j \frac{n_{ij}}{n_{.j}} N_{.j}\right) = \sum_j N_{.j}^2 \operatorname{var} \frac{n_{ij}}{n_{.j}} + \sum_{l \neq m} N_{.l} N_{.m} \operatorname{cov}\left(\frac{n_{il}}{n_{.l}}, \frac{n_{im}}{n_{.m}}\right)$$

$$= \sum N_{.j}^2 \operatorname{var} \frac{n_{ij}}{n_{.j}}$$

$$= N^2 \sum_j p_{ij}(p_{.j} - p_{ij}) E \frac{1}{n_{.j}} \qquad \text{by (16) and (18),}$$

and

$$\operatorname{cov}\left(\frac{n_{ij}}{n_{.j}}, \sum_j \frac{n_{ij}}{n_{.j}} N_{.j}\right) = N_{.j} \operatorname{var} \frac{n_{ij}}{n_{.j}} = \frac{N^2}{N_{.j}} p_{ij}(p_{.j} - p_{ij}) E \frac{1}{n_{.j}}$$

by (16) and (18).

Consequently,

$$\operatorname{var} L_{ij} = \frac{N^2}{n}\left\{p_{ij}(p_{.j} - p_{ij})(1 - 2p_{.j}) E \frac{n}{n_{.j}} + p_{.j}^2 \sum_j p_{ij}(p_{.j} - p_{ij}) E \frac{n}{n_{.j}}\right\}. \tag{20}$$

Consider first the case where the column totals only are known. The above estimate then becomes

$$L_{ij} = (n_{ij}/n_{.j})N_{.j}$$

which is obviously unbiased and adds up to the column totals $N_{.j}$.

To compare between the variance of L_{ij} and that of the inflated estimate I_{ij} we are going to use an upper bound for $E(n/n_{.j})$.

Stephan [6], gives several formulas for upper bounds of $E(1/n_{.j})$. We are going to use here the upper bound

$$u = u_1 + u_2 + u_3 + u_4 + \cdots + (t+1)u_t$$

where

$$u_1 = \frac{1-k}{(n+1)p}, \quad u_r = \frac{(r-1)u_{r-1} - k/r}{(n+r)p} \quad (r > 1), \quad k = npq^n(1-q^n)^{-1}.$$

The following table gives the upper bounds for the expected values of $E(1/n_{.j})$ for $n=1{,}000$, $10{,}000$.

Thus, according to the above upper limit u, the variance of L_{ij} will be less than that of I_{ij} as long as $p_{ij} > (np_{.j}u-1)/(nu-1)$, which

TABLE V

UPPER BOUNDS FOR THE EXPECTED VALUES OF $E(1/n_{.j})$ WHEN $n=1{,}000$, $10{,}000$

$n=1{,}000$		$n=10{,}000$	
$p_{.j}$	Upper bound	$p_{.j}$	Upper bound
1	010091761	1	001000902
.2	005020182	2	.000500200
3	003341156	.3	000333411
.4	.002503765	4	000250038
5	002002005	5	000200020
.6	001667780	.6	000166678
.7	.001429185	.7	.000142863
.8	.001250313	8	.000125003
.9	.001111235	9	000111112
.01	.113007	.01	.010101031
.05	.020395762	.05	.002003815
		.001	112090
		.005	.020414916

is practically always true. When $n=1000$, var L_{ij} would be larger only if the expected frequency in the cell under consideration is less than one individual for any value of $p_{.j}$ as low as .05. When $p_{.j}=.01$ the corresponding expected frequency becomes 1.2 individuals. In the case $n=10{,}000$ the variance of L_{ij} is less than that of I_{ij} as long as the expected cell frequency is not less than one individual, even if $p_{.j}$ is as low as .001. Similarly in the case $n=100{,}000$ the above inequality is true as long as the expected cell frequency is not less than one individual, even if $p_{.j}$ is as low as .0001. In the asymptotic case where $1/n_{.j}=1/np_{.j}$ the variance of L_{ij} will always be less than that of I_{ij}, for any p_{ij}.

The fact that $E(1/n_{.j})$ is asymptotically equal $1/np_{.j}$ can be seen from statistical considerations. Since $L_{ij}=(n_{ij}/n_{.j})N_{.j}$ is identical with the maximum likelihood estimate in this case, the variance of L_{ij} for large n will be identical with that of the maximum likelihood estimate.

Now, the probability element in this case is

$$f = \frac{1}{N}\left\{N_{11}^{X_{11}}N_{21}^{X_{21}}\cdots(N_{.1}-N_{11}-N_{21}\cdots)^{X_{r1}}N_{12}^{X_{12}}\cdots\right\}$$

where $X_{ij}=0$ or 1 and $\sum X_{ij}=1$

$$EX_{ij} = p_{ij} = \frac{N_{ij}}{N} \qquad \begin{cases} i = 1, 2, \cdots, r \\ j = 1, 2, \cdots, s-1. \end{cases}$$

Since

$$\frac{\partial \log f}{\partial N_{ij}} = \frac{X_{ij}}{N_{ij}} - \frac{X_{rj}}{N_{rj}}, \quad -\frac{\partial^2 \log f}{\partial N_{ij}^2} = \frac{X_{ij}}{N_{ij}^2} + \frac{X_{rj}}{N_{rj}^2}, \quad -\frac{\partial^2 \log f}{\partial N_{ij}\partial N_{kj}} = \frac{X_{rj}}{N_{rj}^2},$$

$$-\frac{\partial^2 \log f}{\partial N_{ij}\partial N_{kl}} = 0,$$

it follows that

$$-E\frac{\partial^2 \log f}{\partial N_{ij}^2} = \frac{1}{N^2}\left(\frac{1}{p_{ij}} + \frac{1}{p_{rj}}\right), \quad -E\frac{\partial^2 \log f}{\partial N_{ij}\partial N_{kj}} = \frac{1}{N^2}\,\frac{1}{p_{rj}},$$

$$-E\frac{\partial^2 \log f}{\partial N_{ij}\partial N_{kl}} = 0\,.$$

The covariance matrix of the maximum likelihood estimates of $N_{11}, N_{21}, \cdots$ is

$$\frac{1}{n}\|\sigma^{lm}\| = \frac{1}{n}\|\sigma_{lm}\|^{-1} = \frac{1}{n}\left\| - E\frac{\partial^2 \log f}{\partial t_l \partial t_m}\right\|^{-1}$$

where the t's denote the parameters N_{ij}.

Substituting in

$$\left\| - E\frac{\partial^2 \log f}{\partial t_l \partial t_m}\right\|$$

and calculating its inverse we can easily see that the large sample variance of the maximum likelihood estimate of N_{ij} is $(N^2/n)p_{ij}(p_{.j}-p_{ij})/p_{.j}$ and since this is equal to $N^2 p_{ij}(p_{.j}-p_{ij})E(1/n_{.j})$, it follows that, for large n,

$$E\frac{n}{n_{.j}} \simeq \frac{1}{p_{.j}}\,.$$

Let us consider now the case where both sets of marginal totals are known. In this case we use the original estimate (19) of which the variance is given by (20). We are going to show that, for large n, the variance of L_{ij} is smaller than that of M_{ij}.

Assuming that n is large enough to enable us to put $E(n/n_{.j}) = 1/p_{.j}$ we get the following formula for var L_{ij},

$$\text{var } L_{ij} = \frac{N^2}{n}\left\{\frac{p_{ij}(p_{.j}-p_{ij})}{p_{.j}}(1-2p_{.j}) + p_{.j}p_{ij}(p_{.j}-p_{ij})\right.$$

$$\left. + p_{.j}^2\sum_{k\neq j}\frac{p_{ik}(p_{.k}-p_{ik})}{p_{.k}}\right\}. \tag{21}$$

In order to clear this formula from the p_{ik}'s which do not appear in the variance of M_{ij} we are going to maximize (21) with respect to them. In other words we maximize

$$\sum_k \frac{p_{ik}(p_{.k}-p_{ik})}{p_{.k}} \text{ subject to } \sum_k p_{ik} = p_{i.} - p_{ij}. \tag{22}$$

Differentiating with respect to p_{ik} we get $p_{ik} = \lambda p_{.k} = (p_{i.} - p_{ij}/(1 - p_{.j}))p_{.k}$ (from 22).

Hence max $\sum_k p_{ik}(p_{.k} - p_{ik})/p_{.k} = (p_{i.} - p_{ij})(1 + p_{ij} - p_{i.} - p_{.j})/(1 - p_{.j})$ and consequently the maximum value of var L_{ij} with respect to p_{ik} is

$$\frac{N^2}{n}\left\{p_{ij}(p_{.j}-p_{ij})\left(p_{.j}+\frac{1}{p_{.j}}-2\right)\right.$$
$$\left.+\frac{p_{.j}^2}{1-p_{.j}}\,(p_{i.}-p_{ij})(1+p_{ij}-p_{i.}-p_{.j})\right\}. \qquad (23)$$

Equation (23) would be less than or equal to var M_{ij} when

$$p_{ij}(p_{.j}-p_{ij})\left(p_{.j}+\frac{1}{p_{.j}}-2\right)+\frac{p_{.j}^2}{1-p_{.j}}\,(p_{i.}-p_{ij})(1+p_{ij}-p_{i.}-p_{.j})$$
$$\leqq p_{ij}q_{ij}+2p_{ij}(3p_{i.}p_{.j}-p_{i.}-p_{.j})+p_{i.}p_{.j}(p_{i.}+p_{.j}-4p_{i.}p_{.j})$$

or

$$p_{ij}^2\left(3-p_{.j}-\frac{1}{p_{.j}}-\frac{p_{.j}^2}{1-p_{.j}}\right)+2p_{ij}p_{i.}\left(1-3p_{.j}+\frac{p_{.j}^2}{1-p_{.j}}\right)$$
$$-\,p_{i.}^2p_{.j}\left(1-4p_{.j}+\frac{p_{.j}}{1-p_{.j}}\right)\leqq 0$$

or

$$\frac{p_{ij}^2}{p_{.j}(1-p_{.j})}\,(1-2p_{.j})^2-\frac{2p_{ij}p_{i.}}{(1-p_{.j})}\,(1-2p_{.j})^2+\frac{p_{i.}^2p_{.j}}{1-p_{.j}}\,(1-2p_{.j})^2\geqq 0$$

i e.,

$$\frac{(1-2p_{.j})^2}{p_{.j}(1-p_{.j})}\,(p_{ij}-p_{i.}p_{.j})^2\geqq 0$$

which is always true.

Equation (23) and var M_{ij} will be equal only when a) $p_{ij}=p_{i.}p_{.j}$ or b) $p_{.j}=\frac{1}{2}$. Thus, for large n, var L_{ij} is always less than that of M_{ij}. The variances will be equal only in the following two cases:

a) All the cell probabilities in the row containing the cell under consideration are proportional to their column totals, i e,

$$p_{ij}=p_{i.}p_{.j} \qquad \text{for all } j$$

or

b) $p_{ik}\propto p_{.k}$ for all $k\neq j$ and $p_{.j}=1/2$.

L_{ij} is thus more efficient than M_{ij} and the computational labor is almost the same for the two estimates. The domain in which var L_{ij} is larger than var I_{ij} and which can be defined by the two inequalities

$$p_{i.}+p_{.j}<1$$

and

$$p_{ij} < \frac{p_{.j}}{(2p_{.j}-1)^2}\left(\left[p_{.j}(1-p_{.j})(1-p_{i.})\{p_{.j}(1-p_{.j})+p_{i.}(1-3p_{.j}+p_{.j}^2)\}\right]^{1/2} - p_{.j}(1-p_{.j}-p_{i.}p_{.j})\right)$$

is naturally smaller than the corresponding domain in the case of M_{ij}. For instance, when $N=100{,}000$ and $N_{i.}=N_{.j}=40{,}000$, var $L_{ij}>$ var I_{ij} when the expected cell frequency is 3,592 (4,000 in the case of M_{ij}). If $N_{i.}=N_{.j}=10{,}000$ the corresponding figure will be 390 (470 in the case of M_{ij}) We should not forget, however, that the formula (23) which we have been using for the variance of L_{ij} is actually a maximum and consequently the above figures which correspond to L_{ij} will always be reduced when the true variance (20) is used

REFERENCES

[1] Deming, W. D., and Stephan, F. F., "On a least squares adjustment of a sampled frequency table when the expected marginal totals are known." *Annals of Mathematical Statistics*, 11, No 4 (1940), 427–44.

[2] El-Badry, M. A., *Extension of Markoff's Theorem on Estimation*, Ph.D. Thesis, London, 1948.

[3] Fisher, R. A., *Statistical Methods for Research Workers*. London: Oliver and Boyd. 1925.

[4] Smith, J. H, "Estimation of linear functions of cell properties," *Annals of Mathematical Statistics*, 18, No. 2 (1947), 231–55.

[5] Stephan, F. F., "An iterative method of adjusting sample frequency tables when expected marginal totals are known," *Annals of Mathematical Statistics*, 13, No. 2 (1942), 166–78.

[6] Stephan, F. F., "The expected value and variance of the reciprocal and other negative powers of a positive Bernoullian variate," *Annals of Mathematical Statistics*, 16, No 1 (1945), 50–61.

THE APPLICATION OF SAMPLING PROCEDURES TO BUSINESS OPERATIONS*

HOWARD L JONES
Illinois Bell Telephone Company

The purpose of this short article is to give business executives an insight into sampling theory and procedures without confusing them with mathematical symbols or unexplained technical terms Sampling is defined, and the principal applications in the telephone business are briefly described The procedures used are grouped into three broad categories—judgment sampling, systematic sampling, and random sampling—the relative advantages of which are pointed out Various ways of minimizing the cost of random sampling are discussed. A word of caution is added regarding the dangers of improperly selected samples.

SAMPLING procedures, as a business technique, have now been developed to the point where desired information about a large population can usually be obtained by examining a relatively small sample. Furthermore, the reliability or precision of the sample, as it is called, can be computed from the sample itself.

In this article, the nature and purpose of sampling are described, the different techniques available are discussed, and the advantages of scientific, or random, sampling are emphasized A section is devoted to methods of minimizing the cost of a sample of this type The material is based on telephone company operations; but similar applications occur in many types of business.

WHAT SAMPLING IS

Sampling means selecting and studying a relatively small number of individuals in order to find out something about the population from which they are selected These individuals may be persons, or they may be animate or inanimate things For convenience, they may be referred to as *items* The number of items in the population may be small, or large, or may even be infinite

Sampling procedures are useful in situations where examining all the items in a population and recording and summarizing the results of such examination would be expensive, or time-consuming, or both, and where some degree of approximation in the final results is permissible In many such situations, a sample of a few hundred items, properly

* Revision of paper presented at a luncheon meeting of the Chicago Chapter, American Statistical Association, October 16, 1952

selected, provides information of sufficient accuracy with respect to a population of several million items. The economies obtainable by employing small samples are frequently of considerable magnitude. Moreover, examinations of large numbers of items can result in less accuracy than is produced by a statistical sample properly drawn and observed by small numbers of well-trained and closely supervised personnel. Repetitive operations beyond a certain point produce inspection fatigue which results in errors and thereby adversely affects the results.

PURPOSES OF SAMPLING

When we employ sampling procedures in the telephone business we usually have one of three purposes in mind:

A. To get information that is available but not ordinarily summarized.
B. To supply a framework for making some kind of survey.
C. To provide a basis for verifying the accuracy of the information shown on our records.

An illustration of sampling with the first purpose—*to obtain available information not ordinarily summarized*—is the sampling of our customers' accounts to determine their local telephone usage. For most of our customers in Chicago, there is a tabulating card that shows the monthly usage of local service, and the charge for such service. We need to know this average monthly usage for customers in each rate classification, and the distribution of that usage, that is, the proportion of our customers that use 0–10 messages per month, the proportion using 11–20 messages and so on. It would be very expensive to summarize this information 100 per cent; and so for about 25 years, we have been relying on sampling procedures.

Samples with the second purpose mentioned—*to supply a framework for some kind of survey*—are taken periodically in connection with our attempts to measure the attitude of our customers toward the telephone company, particularly toward the service we furnish and the charges we bill for it Accounting or plant records, and sometimes telephone directories, are used to select the sample customers. These customers are then interviewed, by telephone or otherwise, and asked to give answers to a set of questions. The analysis and summarization of these answers yields the information we want Similar surveys have been taken in the last few years to measure the attitudes of supervisory employees toward the company and higher management.

Another kind of sampling we do occasionally is the sampling of plant and engineering records to provide a framework for appraising the per

cent condition of our telephone property. This condition is one of the factors we ask the regulatory commissions to take into account in fixing the rates we charge for telephone service.

The third purpose of sampling—*to verify the accuracy of the information shown on our records*—is essentially the same as when sampling inspection is employed in a factory to control the quality of the product. This quality control function has two broad aspects:

1. Acceptance sampling.
2. Process control.

Acceptance sampling means the sampling inspection or auditing of work received from another department or company to insure its meeting specifications as to quality. The same techniques can also be applied to testing the quality of work leaving a department or company, to insure its meeting certain standards from the point of view of the customer or the department receiving it. These techniques are particularly well suited to occasional or periodic tests by independent auditors.

Process control means day-to-day routine inspection or verification of the result of an operation in the department where it is performed, to see that the average quality is up to standard, and that the variations in quality are no greater than would be expected from chance causes alone Sampling plans have been found useful for the routine verification of clerical work of experienced operators where extreme accuracy is not required.

Control charts and other quality control techniques[1] are also used in setting objective standard levels for the quality of various operations and in bringing the actual quality up to these levels These charts not only show the comparison between actual performance and the standard level, but they also incorporate control limits, statistically determined, to indicate whether or not differences between actual and standard performance are significant.

TYPES OF SAMPLING PROCEDURES

The commonly used procedures for sampling a population might be grouped into the following broad categories:

A. Judgment sampling.
B. Systematic sampling.
C. Random sampling.

---

[1] Quality control techniques, as applied to factory operations, are described in several textbooks. Among the best known is E. L. Grant's book [3] which is intended as a manual for production and inspection supervisors, engineers, and management.

In *judgment sampling*, the selection of sample items is left to the judgment of some responsible person or persons familiar with the characteristics of the population that is being sampled and therefore presumed to have sufficient knowledge to select representative items. If the characteristic we wish to investigate is closely associated with one or more other characteristics, such as geographical location, and if the distribution of the population with respect to these other characteristics is known, the sample is usually selected so that its distribution by known characteristics will closely agree with the distribution of the population. The sample is then called a *quota sample.*

In *systematic sampling,* all the items in the population are arranged in some kind of sequence, if they are not already so arranged, and every rth item is drawn into the sample. For example, if we wish our sample to include one per cent of the population, we start with some item among the first hundred, and then select every hundredth item thereafter If the first item is selected at random from among the first r items in the population, then every item in the entire population will have the same chance $(1/r)$ of being selected Systematic sampling with a random start may be included with random sampling in the general class of *probability sampling* where every item has a known probability of being drawn into the sample.

In *random sampling,* the actual selection is nearly always determined by assigning a serial number to every item in the population to be sampled and then employing a table of random numbers. Random sampling is also referred to as *scientific sampling.* When a statistician speaks of "sampling" without a qualifying adjective, it is presumed that he refers to this kind of sampling.

There are various subclasses of random sampling. In *pure random sampling,* every selection is made in such manner that all items in the population have a chance, and the same chance, of being selected. In *stratified sampling,* the population is first divided into several parts, or strata, and each selection is made so that only items in a particular stratum have a chance of being drawn into the sample at that particular time In this respect, stratified sampling is like quota sampling. In quota sampling, however, the selection within each stratum is left to the judgment or convenience of the person making the selection; while in stratified sampling, the selection within each stratum is purely random.

Several telephone companies in the Bell System have made considerable use of a sampling procedure consisting of the selection of random subsamples, where each subsample is a sample of the entire population.

This general method of sampling, with the number of subsamples nearly always equal to 10, has been used by W. E. Deming in sampling surveys for Illinois Bell and other Associated Companies to determine the per cent condition of telephone property. It was suggested by J. W. Tukey of Princeton University and the Bell Telephone Laboratories.[8]

This type of sampling is a kind of stratified sampling. To design a plan of this kind, in its simplest form, we first assign a serial number to every item in the population. A counting interval is then computed from the formula

$$k = \frac{10N}{n},$$

where N is the highest serial number in the population, n is the desired number of items in the sample, and k is the required counting interval. Fractions in the quotient are rounded off to some adjacent whole number. To illustrate, let us suppose that the serial numbers 1, 2, 3, $\cdots$ are assigned in order to a population consisting of 99,960 items Suppose we wish to have 500 items in our sample. Then the counting interval might be computed as

$$k = \frac{10 \times 99{,}960}{500} = 1{,}999.2.$$

This could be rounded off to the next higher integer, or 2,000.

The next step is to divide the population into slices, with the first k items in the first slice, the second k items in the second slice, and so on. In other words, if the counting interval were 2,000, we would include items numbered 1 to 2,000 in the first slice, items 2,001 to 4,000 in the second slice, etc

Next we select 10 sample items from each slice For the first slice, the 10 items are selected by employing a table of random numbers. At this point, we may choose either of two further procedures.

One procedure is to add multiples of k to the serial numbers of the sample items in the first slice. Suppose we let $a, b, c, \cdots, j$, represent the random serial numbers of the sample items in the first slice. Then the sample items in the second slice will be those with serial numbers $a+k, b+k, c+k, \cdots, j+k$. The sample items in the third slice will be those with serial numbers $a+2k, b+2k, c+2k, \cdots, j+2k$. After all the sample items have been selected in this manner, they are combined

[8] See [2, pp. 96 and 353].

into ten subsamples, the first subsample consisting of items numbered a, $a+k$, $a+2k$, $\cdots$; the second subsample consisting of items numbered, b, $b+k$, $b+2k$, $\cdots$; the tenth consisting of items numbered j, $j+k$, $j+2k$, $\cdots$. As a result, each of the ten subsamples will include one sample item from each of the slices into which the population has been divided, except for a possible slight variation in the last slice.

An alternative method, frequently used, is like the foregoing in that each subsample contains one item from each slice. But the ten sample items for each slice are selected at random, and not by adding multiples of k to the serial numbers of the sample items in the first slice. These ten sample items are assigned to subsamples in order of random selection.

RELATIVE MERITS OF DIFFERENT SAMPLING PROCEDURES

The best sampling procedure for a particular problem will usually depend on two considerations:

1. The cost of selecting and inspecting the sample and then analyzing the results.
2. The desired reliability of the results.

Let us compare the various classes of procedures from these two standpoints.

Judgment sampling is generally speedy, and when an emergency arises requiring an immediate answer, we may have no practical alternative to this procedure. Or when we want only a very rough idea about some characteristic of a population, the time and added expense of carefully designing and selecting a probability sample may not be warranted. Thus, in the process control of keypunching errors, the selection of consecutive items on a judgment basis seems to have important practical advantages in many situations.

The reliability of a judgment sample depends on the judgment and skill of the person or persons making the selection, and may therefore be very good or very bad. The difficulty is that in most situations there is no way to determine objectively just how reliable the sample actually is. The reliability is purely a matter of opinion. Because of this uncertainty, there is frequently a tendency to have the sample include a fairly large proportion of the population—say, 10 or 20 per cent—with the thought of insuring its reliability. As a result, the sampling procedure may turn out to be costly and time-consuming, without actually overcoming the inherent uncertainties to any appreciable degree. As

mentioned earlier, the fatigue resulting from a large volume of repetitive operations may also result in errors in observing and summarizing. A large sample is not necessarily a good sample; but it is nearly always an expensive sample.[3]

In *systematic sampling*, where the sample consists of every *r*th item in the population, the manner of selection is simple, and easy to explain to the average office worker who will do the sampling. It may therefore be the most practical procedure where a continuous sample must be selected by clerical forces on the basis of written instructions issued by the statistician. This procedure is used, for example, in continuous samples of telephone toll messages.

The precision of the result obtained from systematic sampling can be determined from the sample data themselves, provided we know or can safely assume that there is no relationship between the characteristic of the population we are trying to investigate and the order in which the items in the population are arranged.[4] In many situations, however, an assumption of this kind would not be correct. For example, if we were to study the telephone usage over customers' lines by selecting every hundredth line in telephone number order, the result of the first number and every subsequent number in the sample ended in the digits 00 would be substantially different from the result if the first and every subsequent number ended in the digits 99. The assumption that there is no relationship between the arrangement of the population and some particular characteristic should be avoided, if possible, unless tests have shown this assumption to be reasonably correct.

Random sampling, when the procedure has been carefully designed by a competent statistician, has the unique advantage that the precision of the final results can be determined objectively without making questionable assumptions.[5] In other words, we can compute a precision range and state that if all the items in the population were examined, the result would fall within that range, with the knowledge that statements of this kind will be correct practically every time. This knowledge rests on the theory of probability, which has been derived mathematically and verified again and again through actual experience as a useful theory in practical sampling problems.

Originally, the one principal objection to random sampling was the

---

[3] For further comments, see [2, pp. 9–11].

[4] The precision can also be computed if we know the relationship between the serial numbers and the degree of correlation between the various items in the population; but in many situations, we know even less about this relationship than we know about the property we are investigating by sampling procedures. A good technical discussion of systematic sampling is given in [1, Chapter 8]. See also [4 and 5].

[5] The discussion on "confidence limits" in [1] may be helpful in avoiding questionable assumptions. See also [4].

cost of carefully designing a satisfactory procedure and then carrying out this design in selecting the sample. This was especially true in pure random sampling. This objection has now been largely overcome without impairing the unique advantage of scientific sampling with respect to reliability. The improvements which have been made can be incorporated in a random sampling plan with random subsamples, which also has some further advantages peculiar to itself. In the last two or three years, there have been several instances where a sample of this type has proved to be less expensive than a judgment sample or a systematic sample which had been employed in previous surveys of the same kind.

MINIMIZING THE COST OF RANDOM SAMPLING

The improvements in pure random sampling that have reduced the costs of this sampling procedure might be summarized as follows:

1. Improved numbering methods
2. Sampling without replacement
3. Stratification.
4. Disproportionate sampling.
5. Controlled sampling.
6. Multistage sampling
7. Cluster sampling.
8. Random subsampling.

The method of numbering the population sometimes affects the cost of selecting the sample items. It is necessary, of course, that serial numbers be assigned so that when a particular number is selected by using a table of random numbers, no question can arise as to which item in the population corresponds to that number. But it is not necessary that every serial number actually be assigned to one of these items. To illustrate, suppose we wish to select a sample of residence telephone customers in Chicago. The most convenient procedure is to take a random sample of *all* the telephone numbers in Chicago, and then discard those that are not associated with residence customers Where the same customer has more than one telephone number, we also discard auxiliary numbers not associated with the main service so as to give this customer the same chance of being selected as a customer with just one telephone number. If this method of numbering and selecting is followed, however, it is necessary to observe the precaution of making no substitutions to replace serial numbers that are not associated with items in the population in which we are interested; otherwise, the probability of selection is likely to vary from item to item.

The method of numbering usually creates no special difficulty in determining the sample size. If we wish to have the sample include one per cent of the residence telephone numbers, for example, we design our sample so as to include one per cent of all the telephone numbers. The number of residence telephones actually included in the sample will then be approximately one per cent of the total residence telephones. Any slight discrepancy between the actual size of the sample and the expected size is usually not important. In exceptional situations, where a minimum sample size is necessary, we can oversample by about twice the square root of the desired sample size, and then discard any excess items on a random basis

Sampling without replacement, instead of sampling with replacement, is always employed. In other words, we avoid duplications in selecting the sample Formulas are available to determine the effect on the precision of the over-all mean of the sample [6]

Stratification is important if various strata in the population differ considerably from one another in their average values or variability, or in the unit cost of selecting and processing When the population is divided into slices as previously described, the number of strata is equal to the number of slices, and is therefore large in most situations. To take full advantage of the greater precision which this makes possible, the method of numbering the items in the population should be such that similar items tend to be adjacent to one another. For example, in numbering telephone customers for a sampling survey of some characteristic that may be closely associated with the size of the community, the communities should first be arranged in order of size before assigning serial numbers to subscribers in those communities

Disproportionate sampling is usually advisable if there is a considerable difference between strata with respect to the variability of the characteristic we are investigating, or with respect to the cost of inspecting the items to measure this characteristic or the importance of the items in the final conclusions. Formulas are available to determine the most economical way to apportion the sample over these various strata [7] In some cases, it may be desirable to inspect all the items in some unusual stratum For example, in sampling telephone usage per customer, it is usually desirable to summarize the data 100 per cent for customers with exceptionally high usage.

In controlled sampling, the size of the sample for each stratum or class of items in the population, as well as the size of the total sample,

---

[6] See discussion of the "finite population correction" in various places in [1, 5]

[7] See remarks on "optimum allocation" in 1, 2, 4, and 5]

is rigidly specified in advance. When such specification is made on the basis of exact knowledge of the total number of items in each class, the precision of the sample results will be greater than when the distribution of sample items among different classes is left to chance. On the other hand, there are two advantages in not rigidly specifying the sample sizes by classes of items:

1. If the number of sample items for each class is not rigidly specified, the actual distribution of the sample items may be compared with the known distribution of the population as a check on the sampling procedure. Some such check is always desirable.
2. There are sometimes errors in records purporting to show the actual numbers of items for various classes. By not specifying the sample sizes by classes in advance, the assumption that the records are correct is avoided. Moreover, the existence of actual errors in the records may sometimes be disclosed.

It should also be mentioned that when sampling by the use of counting intervals is employed, the greater precision resulting from specifying the sample sizes by classes will usually be of negligible importance, provided the population has first been arranged according to these classes before serial numbers are assigned.

While controlled sampling is practically never desirable, it is nevertheless often desirable to vary the sample proportions between classes. This variation is sometimes most easily accomplished by varying the counting interval from class to class so that the slices of the population are narrower for classes where a higher sample proportion is indicated. Another procedure is to oversample, using the same counting interval for all classes, and then subsequently discard a fixed proportion of certain classes on a random basis. In a sample of telephone poles, for example, the sample poles were divided between those with cable attachments and those without, and only 25 per cent of those without cable attachments were actually inspected. Thinning the sample in this way may be employed when there is little or no information beforehand as to the relative sizes of various classes, or where it is difficult to separate the population into different classes before assigning serial numbers.

Multistage sampling, as opposed to *single-stage sampling*, is another device for reducing the cost of selecting the sample and making the subsequent inspections. In designing a sample of dwelling units in Chicago, the procedure was, first, to select sample blocks, and then to select a sample of the "dwelling locations" in each block, a dwelling

location being defined so as to include from one to six dwelling units. For example, a dwelling location might be a single dwelling house, or a three-flat building, or all the apartments at a particular street address. All units at a particular dwelling location were included in the sample. The procedure in this particular case would be called *two-stage sampling*, the first stage being the selection of the block, and the second stage the selection of the dwelling locations within the block. Since each dwelling location comprised a group or cluster of one to six dwelling units, the procedure followed in this particular case might also be called *cluster sampling*.

The principal advantage of cluster sampling and various multistage procedures in field surveys is the saving in travel time. A considerable saving may also be effected at times in sampling internal records, if the procedure is such that several items are inspected in each sample drawer or file. There are some disadvantages, however, that need to be considered:

1. Where only the first sampling stage is under the direct supervision of the statistician, and subsequent stages are to be carried out on the basis of verbal or written instructions, there is always the possibility that the instructions will not be understood, or that they will be departed from more or less on the assumption that this is an area where "judgment" may be exercised.
2. With certain exceptions, the results of multistage sampling or cluster sampling are more or less biased. This bias is likely to be important if the number of sample units at the first stage is small, and if at some stage there is a high degree of correlation, positive or negative, between the sample size and the characteristic we are trying to measure.
3. The measurement of the precision of the sample results may be difficult.
4. The averages of individual sample clusters are likely to be misinterpreted. In opinion surveys, for example, if the returns for some particular community consist of 10 questionnaires and all are unfavorable, there will be an intuitive tendency to regard the result as significantly bad, even though just one cluster was involved.

It should also be pointed out that if there is a high degree of correlation within clusters, the sample size required for a given degree of precision with multistage sampling may be considerably greater than for single-stage sampling. For this reason the expense of collecting and sum-

marizing the additional information that is necessary may largely or altogether offset the saving in travel time.

Multistage sampling, or cluster sampling, is sometimes necessary; but the possible disadvantages should be kept in mind.

Random subsamples, which constitute an important feature of many plans now in use in the Bell System, save much time in analyzing the sample results. The results are first summarized for each of the subsamples, and the subsample average is computed. The overall average for the entire sample is the average of the subsample averages. To compute the precision range when 10 subsamples are selected, it is necessary only to add the squares of 10 differences, divide by 10, and extract the square root.[8] Random subsamples are also useful in testing for possible bias in multistage sampling.[9]

CONCLUSIONS

The following is quoted from a manual on sampling issued by the American Telephone and Telegraph Company:

> "Properly used, sampling procedures are among the most powerful and valuable tools at the statistician's command. Improperly or carelessly used, they are among the most dangerous. Particularly careful attention should be paid to the selection of the sample. Too often, perhaps, insufficient consideration has been given to the matter of selection, with the result that badly biased samples have been put forth as reliable simply because the samples were large. This point cannot be overstressed. A sample can never be better than its method of selection."

REFERENCES

[1] Cochran, William G., *Sampling Techniques*. New York: John Wiley & Sons, 1953.

[2] Deming, William Edwards, *Some Theory of Sampling*. New York: John Wiley & Sons, 1950.

[3] Grant, Eugene L., *Statistical Quality Control*. Second Edition, New York: McGraw-Hill, 1952.

[4] Hansen, Morris H., Hurwitz, William N., and Madow, William G., *Sample Survey Methods and Theory*. New York: John Wiley & Sons, 1953.

[5] Yates, Frank, *Sampling Methods for Censuses and Surveys*, Second Edition. London: Charles Griffin & Company, 1953.

---

[8] The precision range, computed as suggested here, is equal to 3 times the estimated standard deviation of the overall sample mean. This measure of precision is convenient in many applications of sampling. The measure itself can be made more precise if independent samples are selected for several groups of items and the sample means and variances are then combined by employing Cochran's [1] formulas (5.1) and (5.7), pages 66 and 69.

[9] The sample average is likely to have some degree of bias unless the number of units to be selected is predetermined for each stage of the sampling procedure. Otherwise, the sample average is the ratio of two random variables: the sum of the sample measurements and the number of units selected. Where random subsampling is employed, this bias can conveniently be investigated by methods suggested in [1] and [4] for finding the bias of a ratio-estimate.

ESTIMATION OF THE BRAZILIAN COFFEE HARVEST BY SAMPLING SURVEY

W. L. Stevens
Universidade de São Paulo

This article describes the results of an investigation, carried out on behalf of the *Instituto Brasileiro de Café*, into the possibilities of estimating the Brazilian coffee crop by means of a permanent sampling survey.

The state of São Paulo already has a general sampling survey for the principal crops. With some modification, this would furnish estimates with adequate precision for coffee. In the states of Minas Gerais, Río de Janeiro and Paraná, there exist registers of property owners suitable for use as frames. From them could be constructed indexes of coffee growers, so that the primary sample would be drawn from this index, with a supplementary sample from the remainder. From data of the pilot survey, an estimate is made of the volume of work which would be needed.

The state of Paraná has recently been surveyed by aerial photography. It was found that this photograph provides a very satisfactory frame for sampling. The state of Espírito Santo possesses no register of property owners, and it is suggested that the sampling survey in that state will have to await a survey by aerial photography.

1. BACKGROUND

How to make reliable estimates of the coffee crop is an urgent problem for Brazil, because Brazilian coffee export is about 50 per cent of world export and about 70 per cent of all Brazilian exports. The problem indeed is more than economic: friendly relations with the United States were recently strained by rocketing retail prices which were, in part, the product of incorrect estimates of the effects of frost.

In only one state of the Federation—in the state of São Paulo—is there in existence a permanent sample survey for the principal agricultural products. In all other states estimates of harvests are of a purely subjective character.

Early in 1954, the writer was asked by João Pacheco e Chaves—then Director of the *Instituto Brasileiro do Café*—to advise on methods of sampling survey for coffee crop estimation in the principal states producing coffee for export (see Table 1)

As no information was available—outside São Paulo—on which to base recommendations, the writer replied that it would be necessary to make an investiagtion in the field, which was accordingly undertaken

TABLE 1

COFFEE EXPORTING STATES: AREAS AND EXPORT, 1953

State	Thousand Sq. Mls.	Million Sacks Of 60 Kg.
São Paulo	91	6.1
Paraná	93	3.2
Minas Gerais	222	3.4
Espírito Santo	27	1.8
Rio de Janeiro	17	0.2
Others	—	0.4
Total		15.1

during 1954, and paid for by the *Instituto Brasileiro do Café*. The detailed report of the investigation was presented in October, 1954. This article discusses only aspects of general interest and it must be understood that the views expressed are purely those of the writer.

2. GENERAL CONSIDERATIONS

In a sampling survey for coffee—or indeed any crop for which the information is supplied by the farmer—the elementary sampling unit is necessarily the farm. True area sampling is not practicable because boundaries of areas marked on a map will usually cut through farms and it is not generally possible to obtain separately, data for the portion of the farm lying inside the marked area.

Since the farm is the ultimate sampling unit, the most useful frame would be an index of farms or more generally, an index of rural properties, since, in a country like Brazil, there are many properties not yet used for farming A true index of rural properties identifies each property on a large scale map showing boundaries of all properties. Such an index possesses the enormous advantage, for sampling purposes, of being, by its very nature, free from omissions and duplication. No such index exists anywhere in Brazil.

As an alternative, one may consider using, as a frame, an index of property *owners*. This is by no means the same as an index of properties, since its freedom from omissions and duplications depends entirely upon the assiduity and competence of the staff which maintains the index. There may be omissions, when an owner fails to register his property. There may also be duplications, when ownership of a property is in dispute and both disputants make a registration or, again, if on the

occasion of transference of ownership, an entry for the new owner is added without the entry of the old owner being deleted.

An index of owners is at a further disadvantage in comparison with an index of properties based on a topographical survey. In the former, if there is much delay in registering a change of ownership, it becomes difficult to identify the property; in the latter, it is not even necessary to record the owner, although it may be convenient to do so. Again, changes in area of a property arising from sale of part of the property or purchase of a part or the whole of a neighboring property are more difficult to recognize if we use an index of owners than if we have an index of properties based on a map.

This means that, if we propose to use as a frame an index of rural property owners, we must satisfy ourselves, not only that it is reasonably free from omissions and duplications, but also that it is being kept up-to-date.

Indexes of property owners will be found whenever there is a tax on rural properties. Of the five coffee-producing states, all except Espírito Santo have a rural property tax levied by the state government. In the state of Paraná, there is also a tax on properties, levied by the counties (*municípios*)—or at least by many of the counties. Accordingly, there exist in that state two independent indexes of property owners.

Although a satisfactory general sampling survey may be based on an index showing nothing more than total area of the property—and in fact that is the basis for the general survey in the state of São Paulo—the survey for coffee will not be efficient unless we have supplementary information on that crop. Consequently, we must investigate the possibility of supplementing the general index of property owners with information of the number of coffee trees or of the production in previous years.

Instead of an index, one may use a map as a frame for sampling properties. One needs however a very good map on a large scale, because it is necessary to:

(i) identify, on the ground, any point or area marked on the map,
(ii) draw on the map the boundaries of any farm visited, or
(iii) as an alternative to (ii), locate on the map a definite point of the farm, such as its most northerly point or the position of the farmhouse.

No maps satisfying these requirements exist in Brasil.

What does exist however, in the state of Paraná, is something even

more useful—an aerial photograph of the whole state, completed in 1953, on the nominal scale of 1/25,000 (varying, in fact, from 1 in 24,000 to 26,000). On this photograph, one can, without difficulty, recognize coffee plantations.

Although it would be possible to use an index of property owners as a frame for sampling in Paraná, it was clearly of very great interest to see what could be done with an aerial photograph, not only for use in Paraná, but also as possibly the only practicable solution in Espírito Santo, where, as has been noted, no index of property owners exists.

3. STATE OF SÃO PAULO

As a general sampling survey for the principal agricultural products has been in operation in São Paulo since 1951, the question here was not to organize an independent survey, but to ensure that the general survey would yield estimates of sufficient precision for coffee

The frame used is an index of rural property owners, which is organized by the "Home Secretariat" (*Secretaria da Fazenda*) for the purpose of levying a tax on rural properties The Secretariat's county tax offices prepare each year, a typewritten alphabetical list of property owners. A copy of the list is sent to the Secretariat of Agriculture where it is reproduced on Hollerith cards. This Hollerith file is the population from which the sample is drawn. The information supplied is name of owner, total area, annual tax (not used) and address.

Where the property is large, its name is a sufficient identification. Small properties often have no name and their addresses are often less complete than they should be. The Home Secretariat has however been asked to instruct county officers to give adequate addresses. Even when the address of a sampled farm is insufficient for identification by the field worker, he can almost always make the identification by consulting the local tax office

The original forms, filled in by the owners, contain further information on the use of the land. This information is unreliable and is not used in the survey. Small discrepancies in total area are common, but the few gross errors have been due to mistakes in transcription.

With regard to the possibility of omissions, the a priori argument, if unscientific, is very strong Registration of the property and regular payment of taxes is, to some extent, evidence of title. It is inconceivable that the owner of a valuable farm will try to keep it out of the register in order to evade a relatively small tax which, in any event, will come home to roost with compound interest added, when the farm is sold or transmitted to an heir. Doubtless there are forgotten and abandoned

properties not on the register, but, as these are unproductive, their omission will not affect estimates of production.

On the contrary, one worries more about duplication. It is known that the same piece of land is sometimes registered by two or more claimants. This can happen only in undeveloped or recently developed counties, and even there, the county tax officer is usually aware of the situation. Whatever the amount of duplication may be—and it is very small—it will necessarily tend to lessen, as new areas are brought into agricultural use and disputes of ownership thereby brought to light.

Changes of ownership are normally reported to the tax office by the new owner However, even if he omits to do this, a copy of the registry of sale will, in due course, be sent in by the notary who registered the sale.

An objective check on omissions is provided by listing the owners of all properties contiguous with some or all of the sampled properties and subsequently trying to identify these in the register. The method has yet to be used extensively in São Paulo, but it was used successfully in the pilot survey in other states. If duplications exist one will, sooner or later, include the same property more than once in the sample. This has never yet happened.

The chief source of error is failure to establish a "one-one correspondence" between properties in the register and properties in the field. A typical case is when a farm is divided formally between the heirs but continues in operation as a single unit It is for this reason, that the total area is an important means of identification. The field-workers are instructed to investigate any large discrepancy between area in register and area declared by the farmer on the occasion of the visit.

Experience over several years has shown that the register of property owners in the state of São Paulo is a satisfactory, if not perfect, frame of reference

The sample, of 1500 farms, is stratified geographically and by total area of property, with class limits, in hectares—

3—10—30—100—300—1000, etc.

The work of visiting the farms is undertaken by the officers of the agricultural advisory service, each of whom has charge of a "region" consisting of from one to five counties. With the intention of spreading the work more or less equitably, the sample is drawn so that a region contributes, on the average 10, and never more than 14, farms. This is not an entirely satisfactory method, because it happens that some parts of the state which have recently greatly expanded production of coffee

and other crops have not received a proportionate increase in numbers of regional officers, with the result that they are under-represented in the sample.

The sampling error for the coffee estimates is about 9 per cent, which is too high to be acceptable.

The Secretariat of Agriculture has, however, begun to organize an index of coffee growers. This will show the number of trees and it can be used as supplementary information for improving the precision of the coffee estimates. If this improvement is not enough, it may become necessary to increase the intensity of sampling in the coffee-growing regions, even if this means that the additional work will have to be done by independent observers.

It may be noted that the *forecasts* of harvest are based entirely on the forecasts made by the farmers There is no possibility, at present, of making more objective forecasts by sampling crops in the field. The regional officers have neither the time nor the training for such work. In the course of time, the analysis of data will make possible a calibration of farmers' forecasts in order to remove biases

As the lines of development for obtaining satisfactory forecasts and estimates in the state of São Paulo are sufficiently clear, we may now turn to the other states.

4. REGISTERS OF PROPERTY OWNERS

The states of Minas Gerais, Rio de Janeiro and Paraná all possess registers of rural property owners, organized for the levying of state taxes. In the first two, the registers are maintained in the taxation offices located in each county, while in the last the organization is centralized in the state capital An objective check was made on the quality of these registers by listing all neighbors of farms visited in the pilot survey and trying subsequently to identify these in the registers.

In the first two states the conclusion was reached that the registers are sufficiently up-to-date and free from omissions and duplications to serve as frames. In Paraná the register is not quite as satisfactory. There appear to be delays in recording changes of ownership, perhaps as a result of the centralization of the work. There are also many spelling errors in names, which can be a nuisance in a file kept in alphabetical order. No doubt these minor deficiencies could be overcome.

Fortunately, however, in the state of Paraná, the *prefeituras* (county governments) levy their own tax on rural properties, with the result that a card index or register of properties, entirely independent of the state register, can be found in the county halls (or at any rate, were found in the half-dozen counties visited in the investigation). It was found that

these county registers are in very good order and are therefore suitable for use as frames.

The state registers all give total area of property. Information about coffee either is not given or, where given, is quite unreliable. It was therefore necessary to investigate the possibility of obtaining this information.

The state of Minas Gerais, in addition to the tax on properties, levies a tax payable directly by the grower, on sale of coffee. For fiscal purposes, the state Finance Secretariat has instructed the local offices to set up and maintain a card index of coffee growers. The blank card provided for this purpose shows that the information sought is name of owner, address of farm, number of trees in production, number not yet producing, annual harvest and individual sales.

It is evident that such an index would be of great value for sampling. Unfortunately, it is not yet in proper working order. In some counties, the cards had not even arrived. At the best, the information was deficient, because the name appearing on the card was often not that of the owner of the farm where the coffee was grown, but of someone else—a share-cropper who grew it, a neighboring farmer who threshed and sold it, or even a merchant who was not a farmer at all. For reasons which will appear, it is essential that each coffee grower in the index of growers should be identified with a property owner in the general register. However, the index of growers has only just been started and there should be no difficulty in getting it properly organized if the local officers are given clear instructions.

In the state of Rio de Janeiro, no index of growers exists An experiment was therefore made, in two counties, to see how one could be organized. A group of leading coffee farmers and merchants was called together in the local tax office and names were read to them from the general register. It was found possible in most cases not only to ascertain whether the farm had coffee but also to obtain an approximate idea of the number of trees. It is evident that an index constructed by such a method may contain numerous omissions but, as will appear later, these will not invalidate the estimates.

In the state of Paraná, some counties show number of trees on their card index and some do not. Where the information is not available, the method tested in Rio de Janiero could doubtless be used.

5 SAMPLING METHODS

Assuming that a card-index of growers has been set up, we may now inquire how it can be used for sampling purposes. The supplementary information can be number of trees, previous year's harvest or average

of a series of years. At first one might think that previous harvest would be more useful than number of trees but an analysis of the data from the pilot survey did not confirm this. In fact, coffee, like other perennial crops, shows strong fluctuations from one year to the next. Possibly an average of several years' harvest would be better, there was hardly sufficient data to examine the point. There are also very strong reasons of a practical nature for using number of trees rather than previous harvest It is usually easier to obtain reliable information on number of trees. For example, a farmer can often say how many trees his neighbors have, when he does not know, or does not like to reveal their last year's harvest. Moreover, number of trees is a fairly stable datum so that it would not matter seriously if the information were not rigorously brought up-to-date every year. It therefore appeared of interest to see what could be done, using number of trees as supplementary information.

Two methods of sampling are available:

(i) selection with probability proportional to number of trees,
(ii) stratification by number of trees, with variable sampling fraction.

The former method, involving as it does a large number of division sums, is not one which recommends itself to the writer Let us pass to the second.

Some 190 farms were visited in the pilot survey. In most cases information was obtained of harvests for three years, that of the last year being often a forecast As the harvests vary greatly from year to year, the data are practically equivalent to that which would be obtained from three independent samples The sample was supplemented by some 30 farms in the state of São Paulo.

The sample was stratified by number of thousand trees in production with the following class limits

1—3—10—30—100—300—1000

It was further stratified geographically according to the amount of data available It was then found that the coefficient of variation inside substrata was stable, with an average value of 0 674. Assuming that the sampling fractions are proportional to the standard deviations, we therefore arrive at the following number of farms for a given sampling standard error in the estimate:

Standard Error	Number
10%	50
5	180
3	500

This, of course, takes no account of the effect of errors in the statement of numbers of trees. Such error will tend to increase the standard errors of the estimates.

On the other hand, it can be shown that the classification used is coarse enough to increase appreciably the standard error. It would be advisable to use the slightly finer classification—

1—2—5—10—20—50—100—200—500—1000

It must not be supposed that the index of growers will be complete; indeed there may be a considerable number of omissions. In any case, even if it were complete, one would still want to estimate the number of new trees, for which purpose one must sample farms not yet producing

It is therefore necessary to sample "non-growers," i e, all rural property owners left in the general register after eliminating those which appear in the index of growers. For these, a sample stratified by total area may conveniently be taken The number requred will depend of course on the completeness of the index of growers. It was suggested that, in the beginning, a sample of "non-growers" at least as big as the sample of growers should be drawn. It is noted however, that the field work is not proportionately increased because those which, in fact, do not carry coffee are visited only once.

6. SAMPLING FROM THE AERIAL PHOTOGRAPH

In view of the great possibilities offered by the excellent aerial survey of Paraná, it is unfortunate that the writer succeeded in obtaining the photograph of only one county. The investigation therefore does not have the generality which one would like.

The simplest way of using an aerial photograph is to select points at random or systematically and visit the farms within which the points fall. The boundaries of the farms are then drawn on the photograph. For each farm in the sample, is calculated the ratio: (production)/(area in the photograph). The average of these ratios, multiplied by the total area of the photograph, then gives the estimate of production of the region covered by the photograph.

Using this method, it was found that the ratio has a coefficient of variation of 1.19, the number of farms needed in the sample being accordingly—

Standard Error	Number
10%	140
5	560
3	1570

This method is not efficient, since it makes no use of the fact that the coffee plantations can be recognized in the photograph.

As an alternative, we can take points at random and select the farm only if the point falls inside a coffee plantation. The denominator of the ratio is then the area of coffee plantations (in the photograph) belonging to the farm and the final multiplier is the total area of coffee plantations in the photograph.

Using this method, it was found that the standard error is the same as for a sample drawn from an index of coffee growers stratified by numbers of trees. This is not surprising, since the supplementary information is of the same nature—area of plantations in photograph instead of number of trees.

The method would, however, after a few years, cease to provide valid estimates because it excludes from the sample, farms which had no coffee when the photograph was taken but which are subsequently planted with coffee. In order to bring these into the sample, it would be necessary to use a combination of the two methods, for example: place a network of points on the photograph and choose all which fall in coffee plantations and, say, one-tenth of those which do not. Then the difficulty is that there does not appear to be any efficient way of making estimates which does not necessitate complicated calculations.

How, then, should the photograph be used? One could make an index of all rural properties with boundaries marked on the photograph, but this would take many years. The solution which appears most satisfactory to the writer is to mark the boundaries of the very large farms and join them by lines following the boundaries of smaller properties, without, however, troubling to survey all the smaller properties. This would divide the whole photograph into areas to serve as sampling units. These would be stratified by area of coffee plantations in the photograph. In the course of time, this information would be substituted by number of trees, so that by the time the photograph had become obsolete with respect to its information about coffee, this information would no longer be used. Since the substitution can take place over many years—except perhaps in new regions which are being opened up very rapidly—the collecting of the information on number of trees can be undertaken during the routine of field work. It will become available in any case for areas in each year's sample, and to these can be added information of all areas contiguous to the sampled areas. In addition, in the course of time, areas containing many properties can be progressively broken down.

By these methods, it would be possible to start the sampling survey fairly soon and to ensure that it continues to operate when the photograph no longer provides valid information on area with coffee.

As to the practicability of the work, the investigation did show that—

(i) it is not difficult to reach, on the ground, any point marked on the photograph

(ii) it is not difficult, though it requires care, to draw on the photograph the boundaries of any farm visited.

7. ESPÍRITO SANTO

There is no register of rural property owners in the state of Espírito Santo. To set up and maintain such a register would be too costly if its only purpose were for coffee estimation, though perhaps it could be considered if the state government decided to make surveys of all principal crops. One may ask whether it would be possible to construct an index of coffee-growers. The problem here would be very much more difficult than in the other states because, in the absence of a general register, it would be necessary to ensure an absolutely complete index of growers.

An experiment was made to see whether such an index could be constructed from information contained in the commercial books of coffee buyers. The result was certainly not encouraging. It would entail a great amount of work with no guarantee that it would yield a complete index. Moreover, as it is not practical politics legally to compel the buyers to open their books to inspection, an uncooperative attitude of one or two would ruin the plan

One can also devise schemes for compelling the coffee grower to register his property, but it is doubtful whether they would prove successful in practice.

The writer came to the conclusion therefore, that the problem will only be solved satisfactorily by first making a survey by aerial photography. This would cost about $50,000 but there is no reason why the whole of this sum should be charged to the coffee survey. The photographs would be used also for making maps: maps now available are very poor.

8. OBJECTIVE FORECASTING AND INTERNAL CONSUMPTION

No suggestion is made, as yet, to make forecasts by sampling the crop in the field. To do this it will be necessary to carry out sampling

experiments over several years at different centers, in order to perfect a method. Subsequently, it will be necessary to train the observers, because estimates will be biased unless the sampling plan is rigorously executed.

Sampling of the crop in the field, even if restricted to a sub-sample of the farms, would greatly increase the work. It is not obvious that it would provide forecasts superior to those made by the farmer, provided these are calibrated, as will be possible when data for several years become available

In order to estimate coffee available for export, it will be be necessary to subtract the internal consumption, about which nothing is yet known. Plans are now being made for a sampling survey to estimate internal consumption.

9 PLAN OF WORK

In order to estimate the cost of the survey, it is necessary to know what standard error can be tolerated. This is a job for the economist rather than the statistician but, in the absence of any definite pronouncement, the writer suggests that it would be reasonable to aim at a standard error of 4 per cent for the three separate estimates of São Paulo, Minas Gerais and Rio de Janeiro taken together, and Paraná, and a standard error of 6 per cent for Espírito Santo. This would give a standard error of 2.3 per cent in the total.

For Minas Gerais and Rio de Janeiro together one may therefore suggest a sample of 320 coffee growers and a sample of 320 "non-growers." Assuming that the index of growers is fairly complete, the number of farms to be visited in the second and subsequent surveys would be not much greater than 320 The opinion here is in favor of a series of separate surveys, each survey to take not more than 3 weeks, rather than a continuous survey all through the season. The pilot survey showed that an average of 4 visits per day should be easily achieved, once the farms have been located. With 4 jeeps, 320 farms would therefore be visited in 20 days The writer expressed the opinion that each jeep should contain 2 persons, not only to ease the strain of driving over long distances but also to keep the work going if one person is ill. However, the counter suggestion is for 5 jeeps each with only one person.

The first survey of each year would take much longer. In each county, it would begin with a revision of the index of growers by comparison with the register in the state tax office. The two samples would then be drawn systematically within classes of number of trees and

total area respectively. These farms would then have to be located and visited. Experience in the pilot survey showed conclusively that on the first visit the observer must be introduced to the farmer by a local person who commands the farmer's confidence: if not, the information supplied will often be false. The first survey may therefore extend over a couple of months and would be carried out during the flowering season September to November. While some sort of forecast could perhaps be make, the information sought would be mainly of age distribution of trees, methods of cultivation, equipment, etc

Forecasts and estimates would be made in three subsequent surveys, before, during and immediately after the harvest. These might have to be supplemented by extra surveys after unexpected climatic conditions, such as a frost.

The amount of work for the state of Paraná would be about the same—perhaps less, as the coffee growing is more concentrated There would however be additional work in the first year in marking out areas on the photographs, if that method were used

10 CONCLUSIONS

The result of the investigation may now be summarized:

(i) Estimates in the state of São Paulo can be obtained through the present general survey, using an index of growers for supplementary information and probably increasing the sample in the coffee growing regions.

(ii) A sampling survey can be established in Minas Gerais and Rio de Janeiro, using as frames an index of coffee growers and the tax registers The index of coffee growers would take only a few months to prepare but might prove unsatisfactory in the first year. It would however become adequate in a year or two.

(iii) The aerial photograph of the state of Paraná can be used as a frame for a sampling survey. A year or maybe more would have to be spent on mounting the photographs and marking areas

(iv) There is no way immediately available for establishing a satisfactory sampling survey in Espírito Santo. The problem will be solved if interested parties can be persuaded to make a survey of the state by aerial photography

(v) The cost of a permanent sampling survey would be small in comparison with the great losses suffered by the farmers as a result of the inadequacy of the present methods of estimating and forecasting.

ON THE RELIABILITY OF RESPONSES SECURED IN SAMPLE SURVEYS*

ROBERT FERBER
University of Illinois

THE reliability with which information is obtained on sample surveys has long been a subject of widespread interest. The ways in which such information may be biased border on the infinite,[1] and it is well recognized—though perhaps not always fully appreciated—that the sampling variation in sample-derived estimates of population characteristics may be negligible relative to the biases introduced at other stages in the operation. As a result of this recognition, the literature on the detection and avoidance of bias in the sampling operation has grown tremendously in recent years.[2]

This study represents, it is hoped, a new contribution to this literature, focussing on an aspect of the subject which, although perhaps one of the greatest transgressors of them all, appears to have received little attention. This is the variability in response of different family members to the same question. In other words, how consistent are the replies of various family members on questions relating to different family characteristics—income, family size, purchases, planned purchases, and others? As will be shown in this paper, the consistency on many questions is very limited, and accordingly serious consideration has to be given to the selection of the respondent, as well as of the family or spending unit, in the great number of economic and marketing surveys designed to elicit family data from one member Suggestions of various means of handling this problem are offered in the final section of this paper, following the presentation of the main results of the study.

THE STUDY

In the latter part of 1951 and the first half of 1952 a randomly selected sample of families in Champaign-Urbana were interviewed in such a way that answers of adult members of the family to the same

* This study was made possible by aid furnished by the Bureau of Economic and Business Research of the University of Illinois. The author would also like to acknowledge the capable assistance provided by Phyllis Barnard, Frances Dotson, Jean Robbins and Raymond Twery at various stages of the operation.

[1] As any attempted breakdown of the main sources of bias readily reveals. See Deming, W. E., "On errors in surveys," *American Sociological Review*, 9 (1944), 359–69.

[2] As is evident from a perusal of bibliographies on the subject in such sources as Parten, Mildred, *Surveys, Polls, and Samples: Practical Procedures*, New York. Harper and Brothers, 1950, and Ferber, Robert, *Statistical Techniques in Market Research*, New York: McGraw-Hill Book Company, 1949.

question were obtained separately and simultaneously. The manner in which this was accomplished involved three steps:

1. A personal letter to the family explaining the study as an experimental one designed to test the reliability of consumer surveys and through this knowledge to help eliminate unnecessary surveys and improve the quality of those that are made. The family was notified of its selection in the sample, was asked for cooperation, and told that a staff member of the Bureau of Economic and Business Research of the University would call for an appointment.
2. A visit by a staff member to the household and further explanation of the survey. If cooperation was secured, an appointment was made for a time when all the adult members would be at home.
3. The interview itself. Each adult family member was given a pencil and a questionnaire form and asked to write in his or her answers to the various questions as the interviewer read them. The interviewer also made sure that there was no "cheating." If through one mischance or another, family members did see the others' answers before writing in their own, the interview was discarded.

The questions asked related to the following subjects:

Nature and characteristics of durable goods purchased during the last six months—who made the purchase, who exerted the most influence in the purchase decision, reason for purchase, and how it was financed.

Durable goods purchase plans for the following six months—who was largely responsible for each plan, when and where the purchase might be made, and contemplated method of financing.

Opinion regarding the relative influence of each family member on the purchase of specific durable goods.

Present economic status.

Economic and political expectations.

Family characteristics—size, ages, occupations, income, presence or absence of family budget.

In this article, we shall examine the consistency of the responses on all of the questions except those relating to opinions of the purchasing influence of the various family members with regard to each other. The latter is a complicated matter in itself and has therefore been made the subject of a different study.[3] Because of this low rate of response

[3] See Ferber, Robert, "On the reliability of purchase influence studies," *Journal of Marketing* January 1955, 225-33.

(and especially because of a 38 per cent refusal rate), the composition of the sample differs from that of the population in some respects. To judge by 1950 Census statistics, it contains disproportionately high proportions of two-member families and of families whose head is engaged in professional work. In addition, the population itself, Champaign-Urbana, is dominated to some extent by the presence of the University of Illinois, and therefore may not typify in many ways other American communities. Hence, generalizations drawn from this study have to be made with great care. It is perhaps best to consider this a case study on this subject although, as will be noted later, many of the results would seem to possess a fair degree of generality.

THE FINDINGS

Inasmuch as family characteristics tend to influence the consistency of replies to the other questions asked, we begin with an examination of the consistency of the replies on the few characteristics of the family where such a check was possible. We shall then consider the consistency of the replies with regard to purchases, purchase plans, present economic status, and economic and political expectations

Family Characteristics. There were three questions on each questionnaire relating to family characteristics to which each family member would be expected to give the same answer, namely, family size, family income, and presence or absence of a family budget.

Before considering the actual findings of this study, a few comments on the nature of the sample would seem pertinent.

The sample members were selected at random from a listing of dwelling units in the Champaign-Urbana area. Many of the original sample families, however, either refused to cooperate or did not cooperate sufficiently to provide completed questionnaires. All in all, a total of about 556 families with two or more adults were approached and 237 completed sets of interviews were obtained, a response rate of 43 per cent.

In the case of family size, there was little difference in the replies received from the same family. Only five differences appeared in the 237 sets of replies, which is probably about as good as one might hope for.

With respect to income, differences were much more frequent, although the magnitude of these differences is difficult to assess because respondents were requested to give the interval in which the family's

income fell rather than the actual figure.[4] The extent of disagreement over the interval in which the family income lies is presented in Table 1 segregated by the number of answers obtained, or questionnaires completed, from each family. Thus, two-answer families are those in which two adults (18 years of age or more) were interviewed in this study; they are not necessarily two-member families, or even two-adult families, for family members under 18 years of age were not interviewed and, in addition, not all of the adult members of every family were always present at these interviews.[5] Three-or-more-answer families are, similarly, those from which at least three interviews were procured. Clearly, the consistency of the replies depends on the number of an-

TABLE 1

EXTENT OF AGREEMENT ON FAMILY INCOME, BY ANSWERS PER FAMILY

Statistic	Two-Answer Families	Three-Or-More-Answer Families	All Families
Proportion of families whose income is given in same interval	71%	44%	67%
Proportion of families whose members disagree by one income interval or more	29%	56%	32%
Proportion of families whose members disagree by more than one interval	3%	15%	5%
Sample size*	182	27	209

* Excluding families in which all but one family member answered "don't know" on the income question

[4] The intervals used were under $1,500, $1,500–2,599, $2,600–4,199; $4,200–6,599, $6,600–10,399 and $10,400 and over. The limits of these intervals come out fairly neatly in terms of both weekly and monthly income equivalents.

[5] The actual distribution of sample families by size and by number of answers is as follows.

Number of answers per family	Family size				
	2	3	4	5 or more	Total
2	90	40	45	28	203
3 or more	—	14	8	12	34
Total	90	54	53	40	237

swers received per family, as well as on family size itself, which is why both means of classification are employed in this study.

As is evident from Table 1, differences regarding the interval in which the family income falls were present in more than a quarter of the two-answer families and more than half of the three-or-more-answer families. Some of these differences may well have been fairly small, particularly where the family income might be near the limit of an interval. On the other hand, fairly substantial differences might have existed which were covered up by the use of income intervals, so that on balance the percentages given in Table 1 probably do not overstate the degree of disagreement existing within the sample families regarding their income.

That the extent of disagreement should increase with the number of answers secured per family is only to be expected in view of the statistic used—a "disagreeing" family being one where at least one member's answer departed from that of the others—and in view of the greater likelihood of disagreement as the size of the family, particularly the number of potential wage earners, increases. It might be noted as a matter of interest that in six of the 27 families for which three or more members reported an income figure, each member mentioned a different interval.[6]

Whether the number of potential income earners influences the consistency of replies to the income question was tested by subdividing the two-answer families by number of adults 14 years of age or over and recomputing the percentages shown in Table 1 for each group. Doing this reveals that the proportions of two-answer families whose members disagree by one income interval or more is larger for the families with more adults, rising up to 40 per cent for families having more than three members 14 years of age or more, but this difference is not significant at the .05 probability level

In other respects, there was little a priori to distinguish families disagreeing on income from other families The disagreeing families were not concentrated in any particular occupation (of head of family) or income group.

As with income, many families could not seem to agree as to whether or not they kept budgets. This question was asked in two forms, namely:

a) Do you have a family budget in the sense of keeping breakdowns of all expenditures by the week or month?

---

[6] To avoid any possible confusion as to the entity for which an income figure was desired, the respondent was asked to state his own income first and then the family income; the two items were specified on the questionnaire; and the interviewers generally reviewed this question as the answer was being inserted.

b) Do you have a family budget in the sense of planning future durable goods expenditures on the basis of some schedule?

Table 2 presents the extent of agreement on these two questions by number of answers per family. In about a third of the families, the answer to this question would have differed according to which family member was interviewed. The answer would have been, in fact, exactly the opposite in most of these cases—more than three-fourths of the disagreements involved situations where one family member answered "yes" and another "no."

TABLE 2

EXTENT OF AGREEMENT ON PRESENCE OF FAMILY BUDGET BY ANSWERS PER FAMILY*

	Two-Answer Families	Three-Or-More-Answer Families	Total
Proportion of families agreeing			
(a) Breakdowns of expenditures	71%	41%	66%
(b) Planning durable-goods purchases	62%	39%	59%
Proportion disagreeing			
(a) Breakdowns of expenditures	29%	59%	33%
(b) Planning durable-goods purchases	38%	61%	41%
Sample size			
(a) Breakdowns of expenditures	195	32	227
(b) Planning durable-goods purchases	165	31	196

* Excludes families with one or more members answering "don't know."

The extent of agreement is lower for larger "answer-families," indicating that unanimity of opinion within a given family regarding this question became progressively less frequent as the number of family members answering the question increased. It does not indicate that consistency declined as family size rose, for a tabulation by this breakdown revealed, if anything, the very opposite (though the trend was not statistically significant). This may well be due, however, to the fact that the answer to the budget questions was much more frequently "no" for the larger families.

None of the socio-economic characteristics on which information was secured appeared to be related to the budget questions. No pronounced differences in consistency were evident by income, by occupa-

tion of head, by age of head of the family, or by number of persons in the family 14 years of age or over.

Purchases Reported. Comparison of the correspondence between durable goods purchases reported by the various family members revealed a degree of consistency that will appear shockingly low to many people but probably not too surprising to those with extensive experience in conducting surveys. For only 16 (8 per cent) of the two-answer families, and for none of the three-or-more-answer families were the same purchases reported by all the respondents.[7] In other words, in only

TABLE 3

EXTENT OF AGREEMENT ON DURABLE GOODS PURCHASES REPORTED BY TWO-ANSWER FAMILIES, BY FAMILY SIZE

(Per Cent of Total Families in Category)

Extent of Agreement	Two-Member Families	Three-Member Families	Four-Or-More-Member Families	Total
Agree on all items.				
Some purchases	9%	5%	0%	5%
No purchases	7	0	0	3
Differ by one item	14	15	10	13
Differ by two items	18	15	15	16
Differ by three items	18	15	33	23
Differ by four or more items	34	50	42	40
Total	100%	100%	100%	100%
Sample size	90	40	73	203

16 cases did the two respondents from the same family list the identical purchases or both state that no such purchases had been made during the period in question; the latter accounted for six of the 16 instances.

The extent of disagreement on actual purchases within the sample families is presented in Table 3 for the two-answer families segregated by family size. (For the three-or-more-answer families, there was at least one respondent in every case but one who differed from the other family members by at least four items In 6 of the 34 cases, no two of

[7] The actual question was, "Has anyone in your household bought any durable goods since last ———— (name of month approximately six months earlier), that is, items similar to those listed on card A?" Card A listed about 50 different durable goods.

the three or more family members were in agreement on all purchases reported.) This table brings out two things. One is that contrary to what one might expect, the disagreements that did occur tended to be more frequently over three, four or more items than over only one or two items. For all the two-answer families combined, in nearly two-thirds of the families characterized by disagreements, the answers differed by three items or more, less than one disagreement out of every seven involved only one item.

The second point is that the extent of disagreement increased rapidly with family size About 16 per cent of respondents in two-member families listed identical purchases, whereas this was true in no case involving a family with four members or more The proportion of families characterized by disagreements over large numbers of items was also higher among families with more members

Family size was not the only characteristic found to be related to consistency of replies on durable goods purchases. Families with heads 50 years of age and over were significantly more consistent in their replies than those with heads under 35 years of age Low-income families were significantly more consistent that those at other income levels. Nearly 47 per cent of families earning less than $2,600 per year agreed on what they purchased during the previous six months as compared with only 6 per cent of families at higher income levels Of course, the former did also report fewer purchases This also may be the main reason why the only occupational difference of any significance was the much higher consistency of replies of families whose head had retired.

It is interesting to note that families whose members disagreed on income or on budget records tended to be less consistent. Of the families whose members gave consistent answers on the income question, slightly over 10 per cent agreed on all durable goods purchases as compared to 2.1 per cent agreement for families whose members differed on the income figure—a highly statistically significant difference The corresponding percentages for families agreeing or disagreeing on the budgeting of expenditures were 9 4 per cent and 4 7 per cent, respectively—statistically significant only at the 10 probability level.

The material presented so far provides a general idea of the extent to which families are likely to disagree within themselves as to purchases reported. What is more important from the viewpoint of sample reliability, however, is the degree of understatement of specific types of purchases likely to result from these disagreements One indication of the extent of this understatement is provided by Table 4. For each of the purchase categories designated, this table indicates what pro-

portion of the different total purchases reported by a respondent were omitted by any member of the same family. Thus, if two members of a family reported a particular purchase and one did not, this was counted as an omission.[8] However, the same omission by two or more family members was counted as only one omission. Hence, the 35 per cent "Total" figure in the first row of the table means that one or more family members neglected to mention a car purchase in 35 per cent

TABLE 4

PROPORTION OF TOTAL REPORTED PURCHASES IN PARTICULAR CATEGORIES NOT REPORTED BY ONE OR MORE FAMILY MEMBERS, BY FAMILY SIZE

Category	Family Size			
	Two	Three	Four or more	Total
Cars and motorcycles	32%	43%	33%	35%
Shoes	60	67	64	63
Clothing	70	71	71	71
Tires and tubes	81	71	76	76
Building materials	63	74	84	74
Furniture and rugs	69	78	80	76
Household furnishings (incl. kitchen utensils)	70	83	82	78
Personal items (mainly jewelry)	69	76	89	80
Major household appliances, incl. radio	31	63	59	52
Minor household appliances	67	47	77	66
Total*	61%	69%	72%	68%

* Includes purchases not classified above.

of the instances in which other family members reported such a purchase—a startling figure indeed.

The other figures in this table are even more startling. More than two-thirds of the purchases made by these families were not mentioned by one or more family members. For several categories 75 per cent or more of the purchases were omitted, the omission percentage generally increasing with family size. In other words, a survey conducted among these families seeking to ascertain major household appliance purchases

[8] Although there is always the possibility that the one(s) omitting the purchase may have been correct, e.g., if the purchase were not actually made during the period mentioned in the question. This possibility is taken up in a later section and, to anticipate the results presented therein, it does not appear to have been of much importance in this study.

over the previous six months would, with sufficient misfortune in the selection of family members for interviewing, have turned up *less than half* the purchases that were actually made. This, of course, is an extreme case, as the likelihood of interviewing all the "wrong" family members in practice is very small. The example points up the fact that the percentages cited in Table 4 represent, in effect, upper limits to the degree of omission that might have occurred on purchases reported rather than expected values.[9] The latter are readily obtained, however, if we are willing to make assumptions regarding which member of a family is interviewed in practice Of the numerous assumptions that are possible, the following five alternatives would seem to include the great majority of consumer purchase surveys:

1. Any male adult member of the family.
2. Any female adult member of the family.
3. Any adult member of the family.
4. The head of the family.
5. The wife of the head of the family or the principal female in the family, i.e , the one who does most of the everyday buying and planning

The proportion of purchases in each category that would not have been picked up by a survey of the sample under study following, in turn, each of the sample-member selection procedures outlined above is given in Table 5. The percentages under selection schemes 4 and 5 are simply the number of items omitted in that category by the particular individuals divided by the total purchases reported in that category. There is some duplication between these two selection procedures inasmuch as female heads of families would be interviewed under both schemes.

In deriving omission percentages for the first two selection procedures, allowance has to be made not only for the omission of purchases but also for the probability of different male or female members being interviewed in families possessing more than one adult member of a given sex. This was done by assigning equal probabilities to the selection of family members in such cases and adjusting the number of purchases omitted accordingly.[10] A similar procedure was followed for

[9] In this respect, however, the percentages are likely to be underestimates because no allowance is made for purchases made by a family during this time but not reported by any family member at all

[10] Thus, if there were 25 families having two male adults, and no purchases of a particular type were omitted in 10 cases, one purchase was omitted by one male in 10 other cases, and one purchase was omitted by both males in five cases, the total number of omissions from the viewpoint of the first selection procedure would be 10. In the first ten cases there would obviously be no omissions. However, in half of the second ten cases, and of course in all of the last five cases, one would expect to encounter omissions.

TABLE 5

PROPORTION OF TOTAL REPORTED PURCHASES IN PARTICULAR CATEGORIES OMITTED UNDER FIVE ALTERNATIVE FAMILY-MEMBER SELECTION PROCEDURES

Category	Any Adult Male	Any Adult Female	Any Adult Member	Head of Family	Wife of Head, or Principal Female
Cars	10%	28%	19%	12%	25%
Shoes	38	24	33	41	24
Clothing	39	30	36	40	32
Tires and tubes	22	54	38	22	50
Building materials	30	41	36	36	44
Furniture and rugs	44	34	39	44	33
Household furnishings	48	37	42	46	35
Personal items (mainly jewelry)	50	39	46	55	41
Major household appliances, incl. radios	31	21	26	31	22
Minor household appliances	40	31	35	35	32
Total*	38%	32%	35%	39%	31%

* Includes purchases not classified above

the third family-member selection procedure. The percentages in Table 5 for the first three selection procedures incorporate this adjustment. It should be noted, however, that no explicit adjustments could be made for possible omissions on the part of adult family members not interviewed, which in effect imputes to them the same pattern of omissions as was found for the respondents

The omission percentages in Table 5 are clearly well below those in Table 4, as would be expected It is also clear that the extent of omission varies considerably by the type of purchase (undoubtedly reflecting in part or in whole the effect of size of purchase as well as type), and that for most types of purchases the extent of omission depends to a large extent on which family member happens to be interviewed. In the case of automobiles, tires and tubes, and building materials, interviews with a male member of the family would have produced substantially more purchases reported—oddly enough, in all of these cases more complete purchase information would have been obtained from interviewing any adult male in the family than from speaking

with the head of the household. For nearly all other durable goods purchase categories, however, more complete purchase data would have been obtained from an adult female member, of the household, with little difference between the completeness of replies of just any adult female member or the principal female member of the household.

Although this test indicates that accuracy of response on consumer purchases depends to a large extent on which family members are interviewed in a particular study, the extent of omissions in any case would seem to many to be shockingly high. Thus, no matter which of the five family-member selection procedures were employed, over 30 per cent of the durable goods purchases of this sample would not have been ascertained, and under some of the selection schemes the number of purchases made and not reported in several categories would have actually exceeded the number reported.

There are, however, at least four extenuating factors. One is the length of the period of reference. Had the purchase question related to, say, one month or two months rather than to six months, many fewer purchases undoubtedly would have been omitted Second is the generality of the question itself. Despite the use of a checklist of durable goods, it is reasonable to assume that had the focus been on a specific commodity, or class of commodities, the rate of omission would have been lower. Third is the absence of any restrictions on the size of the durable goods purchases to be reported. As a general rule, one would suspect that the rate of omissions would be related inversely to the lower dollar limit of permissible purchases reported. Fourth is the possibility that some of the omissions were really correct and that the reported purchases in some cases were in error This is a possibility which is discussed shortly

Characteristics of the purchase. Four items of information were solicited regarding each purchase reported They were:

a. The identity of the family member actually making the purchase,
b. Which family member was most instrumental in getting the purchase to be made,
c. How the purchase was financed,
d. The reason for making the purchase at that particular time.

As is evident from Table 6, the extent of agreement on these questions among the different family members varies considerably from one question to another. It is highest on the method of financing the purchase, about which seven out of every eight families were in complete agreement, and lowest on the reason for the purchase and on which family member was most instrumental in effecting the purchase,

about which slightly more than half of the families were in complete agreement.[11] The latter are, of course, on a more tenuous basis than the others, which would account for the lower degree of consistency in their case.[12] It is nevertheless interesting to note that even in the case of a simple factual question as who made the purchase, different answers would have been secured nearly 40 per cent of the time depending on which family member was interviewed.

TABLE 6

EXTENT OF AGREEMENT ON CHARACTERISTICS OF PURCHASES REPORTED*

(Per Cent of Total Purchases to Which All Family Members Gave the Same Answer)

Characteristic	Two-Answer Families	Three-Or-More-Answer Families	Total Sample
Member making the purchase	63%	41%	61%
Member most instrumental in purchase	57	35	55
Method of financing purchase	88	82	87
Reason for purchase	58	43	57

* Excluding in each case families in which one or more members did not answer or were not sure of the answer.

Validity of the omissions. The analysis thus far has been predicated on the assumption that purchase omissions were in fact errors, that all reports of purchases made during the given period were correct. How valid is such an assumption? Could not some of the purchase reports be erroneous in the sense that those purchases were not made during the period studied? In utilizing survey data to estimate quantities of particular products purchased (usually staples), it is a common characteristic to obtain estimates far in excess of known production data adjusted for inventory changes The reason is primarily a phenomenon

[11] If families in which one or more members answered "don't know" were included and "don't knows" counted as regular answers, the above percentages would be reduced by three to five percentage points.

[12] The wording of these questions was:

"Now, as we all know, the person buying a good is not always the same as the one who is most interested in having the purchase made. In the case of each of the above items, which member or members of the family were most instrumental in getting the purchase to be made?"

"Why were each of these purchases made at that particular time?"

It should be noted that the reasons for the purchase were classified into broad categories—"needed it," "special sale," "just wanted it" or "felt like buying it." Had a finer classification been employed, the extent of agreement on the replies would undoubtedly have been considerably lower

known as "telescoping," meaning a tendency to encompass within the reporting period purchases made at other (invariably earlier) times. One may well inquire, therefore, about the extent to which this phenomenon may account for the purchase discrepancies noted earlier.

To answer this question a supplementary survey of 100 additional randomly-selected families was undertaken during January and February of 1955.[13] Essentially the same approach was used as in the original survey except that only the questions relating to past purchases and to family characteristics were asked and the technique was introduced of letting the family members discuss their answers among themselves after the interview had been completed. This enabled the interviewer to determine not only whether or not particular purchases had been reported correctly, but also the reasons for such omissions as were made.

The supplementary survey yielded a much better response than the original survey, a factor that can probably be attributed to refinement of the initial contact. Of 90 eligible families (10 were one-adult families and hence not eligible), 16 could not be contacted (most had moved or were out of town), 15 refused, and 58—or 64 per cent of the eligibles—were interviewed, all but one being two-adult families. Members of these 58 families reported 277 durable goods purchases, and 164, or 59 per cent, were omitted by one or more members This is a somewhat lower percentage than that shown in Table 4, which may be due in part to the emphasis of this questionnaire on purchases Discussion of the omissions after the interviews indicated the following reasons for them:

	Per Cent of Total Omissions
Lack of knowledge	18%
Forgetfulness	79
Telescoping; purchase not actually made	2
Other[14]	1
	100%

Thus, for this sample in 98 per cent of the cases the purchase had been made and the omission was indeed an error. In view of the fact that there is no evidence to indicate that the replies of the sample in

[13] The possibility that telescoping might exist was overlooked by the author and was only brought to his attention after the original survey had been completed. Since the original sample members had been told they would not be reinterviewed, a supplementary survey seemed the next best alternative.

[14] For example, in one case the purchase of a house was omitted although all arrangements had been completed and the deal closed. The wife knew about it but felt that it was not a purchase in her sense because they had not yet moved in.

this respect would have differed from those of the original sample,[15] it does appear that telescoping was not an important factor in this study. Telescoping may well be of greater significance in the case of items purchased frequently than for items purchased only occasionally, as is true of most durable goods.

It is interesting to note from the preceding tabulation that in nearly a fifth of the omissions the reason was lack of knowledge rather than poor memory. The purchases omitted in these cases were mostly what one would suspect—tires and tubes, personal items, household furnishings. In four instances, however, the situation was encountered of a man and wife making most purchases independently of the other, and knowing of each other's purchases only in the case of such "inescapable" items as cars, houses and furniture.

Purchase plans. Plans, of course, are more nebulous than actual purchases, and one would therefore expect more disagreement within a family with regard to purchase plans than with regard to purchases This is especially so in the present case because no attempt was made to define a purchase plan. Rather the objective was to use the term in much the same sense, and with much the same lack of explicitness, as it has been employed in many buying-intention surveys in the past, and in this way to obtain some idea of the reliability of responses secured on such surveys. The opening question on the subject of purchase plans therefore was phrased as follows: "Are any of you planning to buy any durable goods between now and ———— (a period approximately six months hence)?" Succeeding questions asked the respondent to list what he or she believed each member was planning to buy, when the purchase might be made and how it was to be financed.

The extent of disagreement within two-answer families on purchase plans reported is indicated in Table 7. As with actual purchases, disagreements are clearly widespread, but contrary to the former situation, the family members differed most frequently by only one or two items rather than by many items. It also appears that, contrary to what might have been expected, the extent of disagreement on plans is less than on purchases (compare Tables 3 and 7). This is also true for the three-or-more-answer families.

Both of these phenomena may well be explained, however, by the many fewer plans than purchases that were reported. Thus, the number of plans reported by two-answer families averaged 2.5 as compared with

[15] The two samples are quite similar with regard to family characteristics The greater emphasis placed by this questionnaire on past purchases might possibly have reduced the likelihood of telescoping somewhat (though a bias in the opposite direction would not be unlikely either), but the effect is hardly likely to be substantial

TABLE 7

EXTENT OF AGREEMENT ON DURABLE GOODS PURCHASE PLANS REPORTED BY TWO-ANSWER FAMILIES, BY FAMILY SIZE

(Per Cent of Total Families in Category)

Extent of Agreement	Two-Member Families	Three-Member Families	Four-Or-More-Member Families	Total
Agree on all items.				
Some plans	3%	8%	3%	4%
No plans	32	15	10	21
Differ by one item	26	25	20	24
Differ by two items	22	12	27	22
Differ by three items	4	15	14	10
Differ by four or more items	12	25	26	20
Total	99%*	100%	100%	101%*
Sample size	90	40	73	203

* Differences in totals from 100% due to rounding

an average of 5 2 purchases reported by the same families In addition, a much larger proportion of families reported no plans at all than reported no purchases—20 7 per cent as against 2.9 per cent, respectively. For these reasons, there would obviously be much less room for discrepancies to arise in the case of plans than in the case of purchases

This is not to deny the fact that in a survey of this type the answers secured with respect to durable goods purchase plans would have possessed greater reliability than those secured on actual purchases. With respect to plans, the same answers would have been obtained 25 per cent of the time irrespective of which family member was interviewed (taking the entire sample as a whole), and in 48 per cent of the time the answers would not have differed by more than one item With respect to purchases, however, the same answers would have been obtained less than 8 per cent of the time from any family member, and the answers would have differed by *more* than one item for about four families out of five, depending on which family member was interviewed.

Essentially the same characteristics that were found to influence the extent of agreement on purchases reported were also found to influence

the consistency of replies on durable goods purchase plans. As is evident from Table 7, agreement among the different family members on purchase plans becomes less frequent as the family size rises. Families with heads 50 years of age or over were significantly more consistent in their replies than families with younger heads, as was also true for low-income families (those earning less than $2,600 per year) relative to higher-income families. In both of these cases, however, the families exhibiting the greater consistency—families with older heads and low-

TABLE 8

PROPORTION OF TOTAL REPORTED DURABLES PURCHASE PLANS IN PARTICULAR CATEGORIES NOT REPORTED BY ONE OR MORE FAMILY MEMBERS, BY FAMILY SIZE

	Family Size		
Category	Two	Three or more	Total
Cars	62%*	100%	86%
Shoes	100	84	87
Clothing	84	87	86
Tires and tubes	100*	100*	100
Building materials	100	97	98
Furniture and rugs	95	94	94
Household furnishings (inc. kitchen utensils)	100	91	93
Personal items (mainly jewelry)	91	95	94
Major household appliances inc. radio	91	84	85
Minor household appliances	100	95	96
Total†	90%	90%	90%

* Less than ten plans in category.

† Includes plans not classified above.

income families—also reported fewer purchase plans. As before, families that disagreed on their income or on whether or not they maintained a budget also tended to be less consistent than other families on their purchase plans.

The extent to which families disagreed on purchase plans for specific types of goods is indicated in Table 8. This table corresponds to Table 4 on actual purchases and shows for each category of goods what proportion of the different total purchase plans reported by any respondent were omitted by members of the same family. The picture that this table supplies regarding omission of plans by the family members is

even more striking than that provided by Table 4 on omission of purchases. On the whole, nine plans out of every ten reported were omitted by one or more family members. With regard to purchase plans for such items as tires and tubes, building materials, or minor household appliances, interviewing the "wrong" adults would have turned up almost none of these plans. Even in the case of such a major item as an automobile, an "unlucky" survey of these families would have uncovered only 14 per cent of the purchase plans for this commodity that did exist (and this may well be an overestimate because of plans in existence that may not have been reported by any family member)

TABLE 9

PROPORTION OF TOTAL REPORTED PLANS IN PARTICULAR CATEGORIES OMITTED UNDER FIVE ALTERNATIVE FAMILY-MEMBER SELECTION PROCEDURES

Category	Any Adult Male	Any Adult Female	Any Adult Member	Head of Family	Wife of Head, or Principal Female
Cars	27%	73%	48%	41%	73%
Shoes	52	41	50	52	43
Clothing	41	49	47	48	53
Tires and tubes	32	79	56	29	79
Building materials	46	60	54	48	57
Furniture and rugs	52	45	50	51	44
Household furnishings (inc. kitchen utensils)	57	38	49	58	37
Personal items (mainly jewelry)	55	45	51	59	53
Major household appliances, inc. radio	30	57	44	33	59
Minor household appliances	52	48	52	59	48
Total*	46%	49%	49%	49%	50%

* Includes plans not classified above

As was noted for the corresponding tabulation for actual purchases, the figures in Table 7 represent upper limits of the extent of omission that might have occurred in a one-interview-per-family survey of these families, and not the rate of omission that might have been expected to occur in practice in such a survey Estimates of the expected omissions in actual practice are presented in Table 9 under each of the five

alternative family-member selection procedures outlined in connection with the analysis of the consistency of replies on actual purchases (p.797). As was the case with purchases reported, the percentages shown in Table 9 are considerably below those in Table 8 They are nevertheless substantial and are generally much higher than the corresponding figures for actual purchases (Table 5). In other words, under any of the five family-member selection schemes tested, a much larger proportion of plans would have been omitted than purchases, both relating to periods approximately six months in length. In general, about half of the durable goods purchase plans would not have been reported irrespective of which family member was interviewed (though, of course, not the same half in each case).

Substantial differences are evident in the omission rates for different categories by sex of interviewee In the case of auto purchase plans, interviews with female members of the household would have failed to uncover nearly three-fourths of the plans Larger omission rates also characterized the female members' replies with respect to tires and tubes, building materials, and, oddly enough, major household appliances On the other hand, more purchase plans for shoes, furniture and rugs, household furnishings, personal items and minor household appliances were not mentioned by male family members As for actual purchases, omission rates appeared to vary, with a few exceptions, more by sex than by the member's status in the family.

In examining Table 9, it should not be forgotten that the same qualifications attached to the estimates of the proportions of actual purchases omitted also apply to purchase plans. Had the period of reference been much less than six months, or had the question been focussed more directly on a specific category of goods, the omission percentages would undoubtedly have been considerably lower

Present economic status. Only one question, other than the one on income, was asked on the subject of present economic status. This question was whether the respondent thought that the family was financially better or worse able to buy durable goods at the current time than a year ago. As is evident from Table 10, in nearly half of the families the answer varied depending on which person was interviewed. Even when those families are excluded in which the cause of the discrepancy was one or more members answering "not sure," disagreements on the answer to this question remained in one-third of the families. To what extent these differences may be due to the inherent ambiguity of any general question on "ability to buy" rather than to difference of opinion is as yet an unanswered question.

TABLE 10

EXTENT OF AGREEMENT ON FAMILY ABILITY TO BUY DURABLE GOODS, BY ANSWERS PER FAMILY*

(Proportion of Families in Which All Members Interviewed Gave Same Answer)

Item	Two-Member Families	Three-Or-More-Answer Families	Total
Proportion of families in agreement	56%	44%	55%
Proportion in agreement, excluding families where one or more members answered "not sure"	68	50	67
Sample size*	197	34	227

* Excluding families in which one or more members did not answer the question.

Contrary to the findings with regard to the previous subjects, families whose members disagreed on income or on maintenance of a budget were not more likely to disagree on this matter as well. The same was true for income and family size, the only socio-economic characteristics tested in this regard

Economic and political expectations Three questions on expectations were included in the questionnaire They were:

"In the coming six months, do you expect your family income to be higher, lower, or about the same as in the last six months?"

"Is there anything that you believe will be hard to get during the next six months because of our mobilization program?"

"Do you expect our relations with the Communist nations to get better or worse or not change during the next six months?"

As is evident from Table 11, the extent of family agreement on these questions is no better than on ability to buy durable goods, and in the case of one question—future relations with the Communist nations—appears to be considerably poorer. This is a question concerning which much uncertainty existed at the time the interviews were conducted, in 1951–52, and it is possible that the same question asked in late 1953 or early 1954 would have evoked more agreement within the sample families—if for no other reason than that greater unanimity of opinion appeared to exist at the time regarding the international situation. The reason why lesser disagreement was noted on such a related question as the existence of possible shortages because of the mobilization pro-

TABLE 11

EXTENT OF AGREEMENT ON ECONOMIC AND POLITICAL EXPECTATIONS, BY ANSWERS PER FAMILY*

(Proportion of Families in Which All Members Interviewed Gave Same Answer)

Type of Question	Two-Answer Families	Three-Or-More-Answer Families	Total
Family income expectations: up, down, or about the same	65%	56%	63%
Possible shortages due to mobilization program; yes or no†	67	37	64
Trend of our relations with Communist nations: better, worse, or no change	50	24	47

* Excluding families in which one or more members answered "not sure" or "don't know." Sample sizes are approximately the same as those shown in Table 10, except where indicated to the contrary.

† Based on 85 two-answer families and 8 three-answer families.

gram may well be due to the fact that only two definitive answers to this question were possible as against three for the other.[16]

From the point of view of design of public opinion surveys, these findings have interesting implications. To judge by the data in Table 11, the intra-family correlations on many opinion questions may well be zero, if not negative, so that in such situations greater efficiency will be obtained by interviewing more than one family member rather than by scattering of interviews.

IMPLICATIONS OF THE RESULTS

In essence, the foregoing findings may be synthesized into three general propositions. These are:

1. The degree to which the attitudes and expectations of one member represents those of other family members is not high
2. A particular family member generally does not have complete knowledge concerning the durable goods purchases and purchase plans of other family members over a period of approximately six months. The extent of this phenomenon appears to increase for larger-size families.

[16] In accordance with the findings regarding present economic status, no statistically significant relationship was found to exist between extent of agreement on any of these questions and family size, income, and extent of agreement on income or on maintenance of a budget.

3. In quite a few families, information obtained regarding family status and characteristics will differ according to which family member is interviewed

These propositions can not be considered to be fully established on the basis of this one study, but they represent, at the least, hypotheses worthy of further investigation. The first proposition in particular is on tenuous ground not only because of its being based on the results of one sample but also because of the limited extent to which the questions asked on this subject in the survey are representative of the multitude of questions falling under that heading.

The findings of this study are nevertheless so conclusive in themselves as to cast serious doubts on the reliability of information obtained on many consumer surveys by the conventional method of interviewing one family member or any family member that answers the door. More important, however, is the light provided by this study on how the reliability of information obtained on such surveys might be improved, namely, in the following ways:

1. Do not treat economic or political attitudes and expectations of one family member as representative of those of the family as a whole except if each (adult) member's reply is obtained and all answers coincide. Unless all family members' replies are obtained, each reply is best treated as an individual opinion rather than a collective one.

2. The same is true for family characteristics and status. Such information is best obtained from at least two family members, and preferably all the adult family members.

3. Six months would seem to be too long a period for which to solicit information regarding durable goods purchases or purchase plans. This may well be true even if all the family members are interviewed, although no information bearing directly on this point is obtainable from this study.

4. If, nevertheless, such information for a period covering six months or longer is required, the interview is best conducted with all adult family members present and with the questions focussing on particular goods. In fact, one could well assert that, at a given cost, expenditure of resources for more complete interviews with each family would lend greater over-all validity to results of such surveys than use of the same resources to increase sample size (in terms of number of families or household units).

5. Completeness of response on purchases and purchase plans varies by type of good and depends more on the sex of the (adult) interviewee than on family status. In other words, in a given study where only one

family member is to be interviewed, the primary determinant of which member should be chosen is which sex is likely to be more concerned with the purchase of that particular good or category of goods. To judge from this study, the choice in the case of the goods studied might be as follows:

Male	*Female*
Cars	Shoes
Tires and tubes	Furniture and rugs
Building materials	Household furnishings
Major household appliances (for plans)[17]	Major household appliances (for purchases)[17]
	Minor household appliances

For personal items and clothing, interviews may well have to be interpreted on an individual basis, i.e., not related to the family as a whole. It should also be noted that the above categories are broad ones, and that the choice for a particular item in a category might not necessarily be the same as for the entire category.

To some extent the margin of preference in a particular case will be decreased as the period of reference is reduced If information regarding purchases or plans is desired for as brief a span as one month, the above classification may not possess too much relevance.

Above all, this study points up the possibility that the securing of reliable data at least in consumer purchase studies is much more difficult than has heretofore been supposed. The use of an individual as a spokesman for the family or household unit combined with the often implicit acceptance of the reliability of the replies pertaining even to the individual are major sources of sample bias

[17] This may not be as paradoxical as it may seem at first sight. Evidently, the males may have made the plans, but the females make the most use of these appliances, thereby leaving such purchases fresher in their minds

THE COCHRAN-MOSTELLER-TUKEY REPORT ON THE KINSEY STUDY: A SYMPOSIUM*

ALFRED C. KINSEY AND ASSOCIATES; HERBERT HYMAN AND PAUL B. SHEATSLEY; A. H. HOBBS AND R. D. LAMBERT, NICHOLAS PASTORE AND JACOB GOLDSTEIN; LEWIS M. TERMAN; PAUL WALLIN, W. ALLEN WALLIS; AND WILLIAM G. COCHRAN, FREDERICK MOSTELLER AND JOHN W. TUKEY

EDITOR'S NOTE: *Because of the unusual character of the CMT report, an appraisal officially sponsored by organized scientific bodies (the American Statistical Association and the National Research Council), we are adopting an unusual form of review. Kinsey and his associates have been invited to prepare the main review, each of the six critics of the Kinsey report whose criticisms are considered by CMT has been invited to contribute not more than two pages devoted primarily (but not necessarily exclusively) to indicating whether he accepts the CMT position on his criticisms, and CMT have been invited to submit a final word on all these statements All the statements were circulated among all the participants before publication.* W.A.W

* * *

ALFRED C. KINSEY AND ASSOCIATES, *University of Indiana*

THE CMT report on the statistical handling of the data in our volume on *Sexual Behavior in the Human Male* now appears some eight years after the manuscript for that volume was completed, and some fifteen months after our second volume, *Sexual Behavior in the Human Female*, has come into print We are indebted to Cochran, Mosteller, and Tukey for suggesting some of the modifications which were introduced in handling the data in our second volume We are especially indebted to Harold Dorn and Jerome Cornfield who, as statistical consultants in the preparation of that second volume, helped us develop the forms in which we presented our data, and guided us in making more guarded interpretations of our material.

Most of the suggestions made in the CMT report were distinctly useful, and we utilized a large number of them—specifically something more than a hundred of them—in preparing our second volume We disagree with the CMT report on only one major issue, namely the practicality of obtaining a probability sample in our area of research Because of the sensitive nature of the subject involved, we do not believe that probability sampling is practical in any extensive study of

* A review article on *Statistical Problems of the Kinsey Report*, by William G. Cochran, Frederick Mosteller, and John W. Tukey, with the assistance of W O Jenkins Washington, D C · The American Statistical Association, 1954. Pp x, 338. $5.00.

human sexual behavior which attempts to survey the whole population of a large city, a state, or the United States.

To quote and paraphrase from the statement which we have already published (pp. 25 to 31 in our volume on the female) we may note again that our difficulty stems primarily from the fact that most persons have engaged in sexual activities which are socially taboo and ofttimes legally punishable under the existent laws. Even premarital and extra-marital intercourse are criminal acts in many states, and mouth-genital contacts, homosexual activities, animal contacts, induced abortions, all contacts with persons below the age of 18 or even 21, and still other types of sexual activity are punishable in all or nearly all of the states of the Union. Any public disclosure of such behavior may lead to a considerable loss of social prestige, and (in this day of governmental concern over the private lives of its millions of employees) to a loss of employment, even if it does not lead to criminal action. Consequently few persons discuss such matters with anyone, even including their spouses, their clinical counselors, their religious advisors, or their most intimate friends. Certainly they are not easily persuaded to discuss such matters with an interviewer representing a scientific survey.

Our primary problem, therefore, and the one which few of the reviewers of our work have adequately recognized, has been to discover some process by which we could persuade our subjects to divulge their sexual activities with such fullness and completeness as their memories would allow. To do this, we have had to give prime attention to establishing rapport and developing effective interviewing techniques Only then has it begun to be possible to select the sample with which we wished to work.

Sometimes it has seemed that the requirements of accurate sampling were antagonistic to the requirements for adequate reporting But we have not wanted a representative sample of unreliable answers, and neither have we wanted reliable answers from respondents who represented nobody but themselves For this reason we have substituted, for the usual methods of probability sampling, a method of group sampling through which we have tried to secure representatives of a number of the components of the larger population in which we were interested. By making our approaches to individual subjects through the groups to which they belonged, and by working as long as necessary to establish our reputation in each group, we have been able to develop group acceptances of our project—after which it became possible to persuade the individual members, and in many instances a hundred per cent or nearly a hundred per cent of all of the members of each group, to cooperate.

Obviously, these problems are quite different from those which are involved in a study of the incidence of an insect infestation in a cornfield, or the frequency with which a given population attends moving picture performances, or the number of tires which the average householder owns for his automobile. Even economic, social and political surveys rarely deal with data which are as sensitive as those with which we have had to deal. We feel certain that an approach to the lone individuals who might have been the pre-selected respondents in a probability sample (in a study of any large population) would have led to such refusal rates that the sample would have been worthless. Moreover, even those who would have agreed to talk to us would not have given such full and complete information as we have secured through our method of group sampling. This appears to be the explanation of the fact that some of the statistically best planned studies on even less sensitive aspects of human sexual behavior have arrived at generalizations which are completely remote from the experience of clinicians, trained social workers, and all of the case-history studies. Their error obviously stems from the fact that they have given prime attention to their sampling patterns and inadequate attention to the problems of interviewing.

To suggest, as some persons have, that our experience should now make it possible for *us* to secure a good probability sample, indicates some failure to comprehend the especial difficulties which we have faced and still face. Even in such closely knit communities as church memberships, college classes, city rooming houses, prison groups, and fraternal organizations, it has usually taken months or even years to establish the community interest which has finally made random sampling possible. For instance, we worked for four years in the California State penal institutions before we were able to lower the refusal rates on a pre-selected sample to the 5 or 6 per cent rates at which they have stabilized within the last two years.

There is no major area of human physiology, psychology, or psychiatry in which our knowledge has been less adequate than in the area of sexual behavior. For this reason, we have never intended to confine our research to a study of the incidences or the frequencies of the various types of behavior in these United States. From the beginning we have attempted to secure data on the anatomy and physiology of sexual response (Chapters 14,15, and 17 in our volume on the female), on the relation of hormones to sexual response, on the factors which account for particular patterns of sexual behavior, on the significance of each type of behavior in the social organization, on the problem which society faces in attempting to control the social behavior

of its individual members (sex law and sex offenders), and on still other matters. Not more than 20 per cent of the information which we have so far gathered has been brought together in the two volumes which are now published. Data which we already have will be the subject matter of later publications on sex law, on sex offenders, on juvenile sexual delinquency, on prostitution, on transvestism, on homosexuality, on the relation of drug addiction to sexual activities, and on still other matters. In all of these areas we should be able to add to our knowledge, but in every one of these areas other investigators will have to make the more extended studies which we, in one life time, shall not be able to make. We have chosen, in view of our present lack of knowledge on most of these matters, to make a general survey of the whole area of human sexual behavior

The present discussion of the statistical problems which we have faced may contribute materially to the effectiveness of our own further work, and to the planning and execution of the research which other groups may choose to undertake But these discussions will have somehow failed if they do not point up the difficulties which are sometimes involved in applying statistical ideals to particular areas of research.

* * *

HERBERT HYMAN, *Columbia University*, and PAUL SHEATSLEY, *National Opinion Research Center*[1]

THE Editor has asked us to comment on the monograph by Cochran, Mosteller, and Tukey on *Statistical Problems of the Kinsey Report;* more particularly to express our reactions to their treatment of our earlier review of Kinsey. On this latter score, our comment will be brief. We find nothing to criticize in their remarks about our review To be singled out for such detailed attention and occasional criticism is to feel honored and to profit. On the larger treatment of Kinsey by these writers, we have elsewhere[2] remarked in print on the soundness of their criticism and the balance and wisdom they exhibit Having attempted such a review ourselves makes us sensitive to the great difficulties involved. Cochran, Mosteller, and Tukey have done a major work, far exceeding any man's ordinary conception of criticism or review. We should like to address our remaining remarks to the larger values of their work.

[1] "The Kinsey Report and survey methodology," *International Journal of Opinion and Attitude Research*, Vol. 2, No 2 (1948), pp. 183–95

[2] Hyman, Herbert H, "Sexual behavior in the human female—a special review," *Psychological Bulletin*, 51 (1954), p 419

In one section of the present monograph, six previous reviews are examined and the comments of the reviewers are grouped under headings corresponding to the major phases of the research process, e.g., sampling, interviewing, etc. This content analysis, in a sense, provides a rare, perhaps unique, opportunity to see what elements of subjectivity or idiosyncrasy characterize present-day technical standards in research and to what extent agreement exists on standards. If evaluation were capricious and if standards varied, it would provide implicit evidence that no investigator would know how to proceed. Principles for the conduct of inquiries in the behavioral sciences would be lacking. In our opinion, this content analysis of the six earlier reviews, plus the parallel treatment by a seventh, Cochran, *et al.*, establishes that there *is* a body of governing principles. Seven expert witnesses are in fundamental agreement. But it will also be clear to the reader that in some places our present principles fall short. The witnesses don't always agree. The principles either provide no clear standard as to correct practice or they point to a technical problem demanding solution, without being able to point the direction of that solution, given the exigencies of field study of a complex problem. To know the limits of current knowledge is of value in its own right. But here is where Cochran, Mosteller, and Tukey provide another valuable lesson. In their suggestions for solution to some of these problems they instruct the student in that blend of principles and modifications, in that simultaneous orientation to ideals and realities, in that tentativeness rather than dogmatism, which mark the best research practice.

Apart from these larger considerations, there is much more to be noted in this monograph which is of general value. For example, Appendix G on *Principles of Sampling*, published separately in this *Journal* (vol. 49, 1954, pp. 13–35), can be read apart from the study of Kinsey and constitutes a brief, clear and valuable general reference.

The magnitude of this critical work is appropriate for the importance of Kinsey's own work. But Kinsey apart, a documented exposition of principles of research in the context of the case study of *any* research inquiry provides an ideal vehicle for training. This is such an exposition.

For these and other reasons, Cochran, Mosteller, and Tukey are to be commended on the conduct of their committee inquiry and their monographic report.

* * *

A. H. HOBBS AND R. D. LAMBERT, *University of Pennsylvania*[1]

WITHIN the important but restricted territory encompassed by their commission from the National Research Council to evaluate the statistical methodology of the first Kinsey Report, Cochran, Mosteller, and Tukey have carefully surveyed an uneven terrain and posted appropriate warning signs around the principal pitfalls. Since the committee which requested this evaluation also "... wished to make constructive advice available to Dr. Kinsey's group," CMT designed a blueprint which might enable KPM to stay out of some of their methodological dead-end streets in future volumes.

In an interesting commentary upon the scientific status of such investigations of human behavior, CMT acknowledge that a judicious selection of topics from the six critical reviews which constitute the groundwork for their analysis would have enabled them to write "... two factually correct reports, one of which would leave the impression with the reader that KPM's work was of the highest quality, the other that the work was of poor quality and that the major issues were evaded." It is our impression that the authors were notably successful in their stated attempt to avoid both of these extremes. In an over-all view their evaluation is comprehensive in technical coverage, judicious in interpretation, and considerate of all concerned in its critical commentary. Regrettably, but understandably, the nature of their commission permits only brief and incidental mention of criticisms which others regard as trenchant and produces an analysis which, by weighing items separately and on specifically statistical scales, fails to measure the cumulative weight of the KPM conclusions

Though some aspects of the critical interpretations are questioned by CMT, they and the critics seem to agree in substance that a grave omission was involved in KPM's failure to include basic data to permit readers to know the actual number of cases in each subgroup and in cross-groupings, and to enable other investigators to evaluate the conclusions; that failure to specify the questions used in interviews was a more understandable but still serious handicap in attempts to evaluate; that the use of volunteers and contact men almost certainly distorted the sample, as evidenced by large disproportions within the sample and between the sample and the unsampled white male population; that KPM, therefore, had no warrant to project their sample to the entire U. S. white male population; that in view of the obviously non-

[1] "An evaluation of 'Sexual Behavior in the Human Male'," *American Journal of Psychiatry*, 104 (1948), pp. 758–64.

representative sample, too many sweeping generalizations were based on small numbers of cases; that the checks on the sample and on the procedures were inadequate; that the meaning of many tables is obscure, with the number of actual cases in various subgroupings undefined and shifting. That textual statements sometimes do not conform with tabular data. That the accumulative incidence technique requires a degree of stability of behavior and representativeness in the sample which is not present; and that conclusions based on comparisons of successive "generations" are questionable. Thus, in substance, CMT agree with most of the crucial statistical limitations emphasized by the critics.

Despite their objectivity and their eminent fairness to both KPM and the critics, the authors, restricted as they were, are compelled to scan a wide area through a narrow lens. Most of the critics, ourselves included, tried to view Kinsey's first volume from a perspective of erroneous impressions which were likely to be conveyed to its wide and statistically naive audience, even though based upon technically acceptable procedures. Thus when our statements about the misleading nature of the accumulative incidence technique are questioned, and when CMT contend that our criticisms of the active incidence data lack relevance (p. 127) the disagreement, we believe, arises primarily from the differing perspectives through which these procedures and the conclusions based upon them are viewed.

Our original statement relating to this issue reads: "To ascertain the number in any age group who are engaging in the specific activity, use of the accumulative incidence technique would presumably necessitate a high degree of representativeness for all significant factors within the given age group . . ." CMT commented that "these criticisms might be relevant if KPM had used accumulative incidence as if it were equivalent to active incidence." What we meant to refer to was not the accumulative incidence, but the distribution of a few cases over a number of age groups in both the accumulative and active incidence. KPM used recall data to place an individual in each age group younger than the one into which he falls at the time of interview. Where the upper age groups are not representative for important background characteristics correlated with the dependent variable, bias is introduced through each of the lower age groups. When we pointed out that the upper age groups were highly unrepresentative on marital status which is associated with variations in sexual practice, CMT answered that KPM were committed only to relatively equal groups which did not require representativeness within each category. This

we would agree with if the cases from one age group were not used in preceding age groups, thus removing independence.

To indicate what we have in mind, we will work through part of an active incidence table (p 652) which gives the percentages of single males by age and educational group that are in each of the homosexual rating groups. On KPM's scale 4+ includes those who are more homosexual than heterosexual in their contacts and/or psychic reactions, and we will confine our example to them. (It might be added that this table could hardly be more exasperating to anyone interested in checking the computations. The N's are erratic, the totals do not total, the age groups differ from one educational level to another, the matching 9–12 educational level for married males is omitted, and when percentages in each homosexuality rating are applied to the total number of cases whole numbers do not emerge.)

Since KPM fail to give the number of cases in any age group at the time of interview, let us assume for the purposes of illustration that those who report themselves as 4+ homosexuals at the upper age group fell under the same rating in the earlier age groups It would appear that this assumption is not too unreasonable, but even if it were unlikely, the possible bias explained below would only become more variable. Consider the following table computed from KPM's table on page 652 for the single males of educational level 13+.

ACTIVE INCIDENCE OF SINGLE 4+ HOMOSEXUALS BY AGE FOR EDUCATIONAL GROUP 13+

Age	(1) No. of Cases	(2) Cases in x_2-x_1 ..	(3) No. of 4+ Homos	(4) Homos x_2-x_1 .	(5) $\frac{4}{2}$	(6) $\frac{\text{No. 4+ Homos. at age 35}}{\text{4+ Homos. at a given age}}$
15	2846	540	273	135	25.0	8.1%
20	2306	1619	138	56	3.4	15 9
25	687	508	82	37	7 3	26.8
30	179	108	45	23	22.2	48 9
35	71	71	22	21	31.0	100.0

x = number of cases at a given age.

It will be apparent that those in age 15 who do not appear in age 20 (column 2) are divided into two groups; some were interviewed between the ages 15 and 20 and some have passed into the married group and thus at age 20 will no longer be in the sample of single males. Since KPM give no breakdown of interviewees at age of interview, it is impossible to estimate the per cent of the "lost" cases that are due to each

factor. Even if these data were given, a proper analysis would require that the ratios of 4+ homosexuality in each of the two categories of "lost" cases be given With these unknowns it is impossible to estimate the contribution of single males at older ages to the per cent of 4+ homosexuals at the earlier ages.

We can, however, calculate the per cent of the total number of 4+ homosexuals at each age which are contributed by the oldest age group (column 6) under the assumption of stability of rating from the earliest to the oldest ages. Some measure of the significance of this contribution perhaps in terms of the mean square error of each age group would have to be devised, but it is greatest where the cases in a given educational level at the time of interview are concentrated at the older ages and/or where the total number of cases is small. A similar calculation for the educational level 9–12 where the cases are fewer will indicate an even greater contribution from the older age groups.

ACTIVE INCIDENCE OF SINGLE 4+ HOMOSEXUALS, EDUCATIONAL LEVEL 9–12

Age	No of Cases	No of Homos	Age 30, 4+ Homos. / 4+ Homos at given age
15	629	109	22 9%
20	350	70	35 7
25	140	43	58 1
30	67	25	100 0
35 (not given)			

x = Cases at a given age.

There is nothing illogical about the position that those predominantly homosexual should stay single, or to put it in KPM's terms that the group age 35 and still single at the time of interview would have been more likely to be predominantly homosexual at age 15, 20, 25, and 30. The problem is that the older age groups (married and single) are a part of each of the younger age groups, and if they are disproportionate on an important variable such as marital status, then they should not be thrown into the lower age groups with an implicit weighting of one If there is an undue proportion of single males at the 35 age level, then they should be weighted on some such pattern as the U. S. corrections for the relevant universe before they are thrown in with cases interviewed at younger ages. The relatively equal sub-samples are useful for

representing frequencies or proportions in a given cell, and with adequate corrections as a contribution to a weighted mean. It seems to us to introduce dangers of bias when a nonproportional sample from an older age group is used to fill out younger age cells. To dramatize the problem, consider the hypothetical case where 95 per cent of those interviewed at age 35 were single. If enough cases were involved, presumable correction factors could make this a reasonable estimate for homosexuality at age 35. But when these cases appear in the 20 year age group, they are corrected only in terms of the proportion which are single at twenty in the relevant universe. The unusual character of these cases is lost when they are thrown into the twenty year group. Thus we argue that a system of weighting must be developed which will take into account not only disproportions of background factors in the specific age group in question, but also where the individual fits at the time of interview if he is to be used in other cells for earlier ages. Even if the assumption of stability of rating is to be abandoned (and such data should already be available to KPM) weighting for background factors at the time of interview must be used at earlier age groups. In the case used in the example, the bias appears to be upward, but in other sampling distributions it could underestimate the incidence.

After agreeing that the KPM statements relating to homosexuality (p. 665) are carelessly worded, CMT object (p. 145) on similar grounds of careless wording to our statement that KPM's references to high rates of "homosexuality" ". . . refer to an activity which may have occurred no more than once during a lifetime. . . ." Here, again, perspective is an important factor in the interpretation. It was (and is) our impression that KPM rationalized to build up the rates and employed the term "homosexuality" in a manner likely to give an erroneously extreme impression of the magnitude of this form of behavior. On such bases, and being limited to the confines of an article for a non-statistical journal, we felt justified in employing a common rhetorical device as an antidote to the extreme impression conveyed by KPM. Admittedly, our statement is strong, but it is not careless, and on the bases of the ambiguous definitions and of the absence of crucial data which are withheld on the topic we believe that it can be hypothetically rationalized in the same manner as KPM's reiterated and emphasized equally strong statements of opposite tenor to which it was offered as a counter-play.

Persons are rated as 1's on KPM's scale of "homosexuality"

> . . . if they have only incidental homosexual contacts which have involved physical *or psychic* response, or incidental psychic responses *without physical*

> contact. The great preponderance of their socio-sexual experience and reactions is directed toward individuals of the opposite sex. Such homosexual experiences as these individuals have *may occur only a single time or two* . . (pp. 639 and 641, emphasis added).

Definitions of 2's and 3's also include persons who have had either physical *or psychic* responses; either actual experience, or "reactions." Even the 4's, designated as being "predominantly homosexual," are defined as those who " . . have more overt activity and/or psychic reactions in the homosexual. . ." So long as the definitions include those with psychic reactions and those with only incidental experience, the failure of KPM to include crucial definitive data leaves open the possibility that actual homosexual behavior may have occurred no more than once even though the percentages are based on a three-year interval. In the case of some, indeed, it is quite possible that homosexual activities were *never* practiced. (See KPM, p 623.)

Further explanation of points on which CMT question our criticisms would merely reiterate the obvious fact that they viewed KPM from a different angle than did we.

CMT's recommendations for future volumes are, of course, excellent. We agree that a probability sample, even a small one, is urgently needed, and that the per cent and characteristics of refusals is crucial. Aside from the question of taxonomically describing the behavior of some universe, for instance the white males in the U. S., the data could be used to make comparisons among variables within the sample. To this end, CMT's suggestions for merging the background variables into a few composites with high internal correlation and the construction of additive scales not anchored to zero points is a significant contribution. In KPM the clinical tables as currently presented are almost useless. An individual might find that his "normal" incidence varies according to the background characteristic he used to locate himself in the table. Composite variables would make the classifications more useful in both individual and group comparisons.

CMT's formulas which are intended to correct sample estimates for cluster and to estimate optimal adjustment for background characteristics are important ventures. Their use seems limited, as they themselves have noted, in so improbable a non-probability sample as KPM's. We would welcome an example working through these techniques for a cell or group of cells to give us some idea of the magnitude of possible bias involved. We wonder, however, if KPM's method of expanding the cases does not call for some longitudinal as well as cross-sectional estimates of error. Perhaps CMT had in mind using the cases only in the cells appropriate to the time of interview.

As a contribution to the general theory of error in non-probability sampling, the appendices are most rewarding. Within the limits imposed by their commission, the authors have performed a real service, not only in combining and evaluating criticisms of the first volume by KPM, but in developing and describing theoretically imaginative and interesting suggestions for other investigators as well.

* * *

NICHOLAS PASTORE, *Queens College*,[1] and JACOB GOLDSTEIN, *New School for Social Research*[1]

WE ARE impressed with the judiciousness of CMT's appraisal both of the KPM volume and of the comments made by the various critics. We shall confine ourselves to a few constructive points.

In line with Kinsey's biological orientation and previous work in the field of biology, KPM avowedly accept the "taxonomic" approach in their assessment of the sexual behavior of the American population. Whether a biological orientation necessarily leads to such an approach is, of course, another question. In any case, we think that a taxonomic approach, which we identify with the classification of behavior (what *is* the actual behavior of Americans) without including propositions concerning a causal analysis of data, as inherently self-limiting. Since considerable theoretical discussions and empirical work on sexual behavior had preceded KPM, they could have tapped this source for systematically formulating a point of view in advance of their own investigation. Such a systematization would enable the investigator to decide and define the character of the critical variables, the items to be selected for measuring the key variable (in this case, sexual behavior), and the method of statistical analysis in so far as the selection of the variables is concerned. KPM write that the number of theoretic groups in the 12 way breakdown they presented is "nearly two billion." A systematic formulation of problems or hypotheses at the outset would in itself considerably restrict the number of breakdowns of interest (quite apart from the availability of subjects in the various cells). We are informed that 300 items "have been explored" in each history. Yet only a restricted portion of the accumulated data is used (see pp. 63 ff. of KPM for examples). Why collect such data? Again, aside from realistic considerations (amount of time is finite and we can't be interested in everything), a definition of the problem in advance would

[1] "Sexual behavior of the American male: a special review of the Kinsey report," *The Journal of Psychology*, 26 (1948), pp. 377–62.

have helped to determine the choice of interview items Reduction in the number of items would have enabled KPM to explore other facets of behavior which would have served to make the results more usable. Information as to the psychological significance of an activity would be important.

The taxonomic approach forces the investigators to use ad hoc methods in analyzing data. Perhaps the investigators try one method of analysis or another until something of interest comes up. We remain uninformed about attempted and rejected analyses from the potentially large pool of analyses. In Vol. I we find a discussion of the masturbation habits of members of professional classes. In Vol. II there is no corresponding discussion. Again, in Vol I we find considerable discussion of vertical mobility as related to adolescent sexual practices; there is no discussion whatever on vertical mobility in Vol. II. Why the differences in approach?

(1) Our point in sharply drawing a distinction between an actual incidence figure and an accumulative incidence figure (CMT, p 124) lay in what seemed to be a confusion between the two—both in the minds of whose who quoted KPM and in KPM themselves. It seemed especially important to call attention to the distinction since the accumulative incidence figure is a larger number than the actual incidence figure We called the 43.3% figure an "artifact" because it could be taken to characterize the behavior of 2816 individuals involved whereas, in actuality, it characterizes only those 97 individuals who reached 50 years of age We then quoted a passage from KPM (CMT, p 124) which, to us, seemed to illustrate the confusion. In retrospect, we realize now that we may not have been entirely fair to KPM. Nevertheless, we feel that their unclear wording on the point has contributed to a popular misunderstanding For another example (since CMT question our cited example), on p 663, Vol. I, KPM write: "Inasmuch as our present data indicate that more than a third . of the white males in *any population* . *have had* at least some homosexual experience . " (italics ours). Other examples can be found in Vol. I and in Vol II. Further, we are still uncertain as to the value of presenting data in terms of "accumulative incidence figures." Apart from whether the underlying assumption is met, such a figure removes us from the raw data themselves We would like to know (and we may be expressing a personal preference) what the actual incidence figures are, prior to the computation of the "accumulative incidence figures." As a suggestion data can be presented in the form of a scattergram (for certain critical variables at least). The abscissa can refer to the actual

age of interviewee at the time of interview; the ordinate can refer to the reported age of first experience.

(2) We still maintain that KPM incorrectly generalize from their data (CMT, p. 148). KPM assert a generalization and cite, so it would seem, "parental classes 3, 4, or 6" by way of illustration. If KPM wished to exclude those classes which contradicted their generalization, they should have stated so explicitly without leaving it up to the reader to discover this for himself by checking through various tables. Moreover, the reader must discover for himself that "mean frequency" is the index intended by KPM rather than any of the other indices.

(3) CMT's citation (p. 28) of KPM's explanation of different totals in different tables troubles us. On page 208, Table 41 of KPM two sets of mutually exclusive categories are presented. We note that the sum of frequencies for *occupational class* exceeds those for *educational level* by more than 800.

(4) With regard to the existence of a seventh outlet (CMT, p. 228) we were influenced by KPM's wording—"the six chief sources of orgasm" (Vol I, p 193). We thought that they had fetishism and other perversions in mind.

(5) CMT write, with regard to their interviews by KPM, that they made "rough and ready guesses" in so far as frequency figures were concerned Their impression coincides with that of one of the present writers who was interviewed by one of the KPM team in 1948. He felt at that time that frequency figures, especially when they pertained to some practice engaged in 15 or 20 years prior to the interview, were within a given range, sheer guesses. Perhaps similar considerations apply to others who were interviewed. We wonder what the mean and median frequency figures in the KPM tables really signify.

* * *

LEWIS M. TERMAN, *Stanford University*[1]

I WOULD now make the following changes in the quotations by CMT from my review of Kinsey's *Sexual Behavior in the Human Male*. Each quotation is identified by the page where it appears in the volume by CMT.

Page 69. In the fourth line from the bottom of the quotation change the word "reliably" to "appreciably."

Page 95. Leave as is to the third word in line 7. The remainder of this quotation should be changed to read as follows:

[1] "Kinsey's 'Sexual Behavior in the Human Male': some comments and criticisms," *Psychological Bulletin*, 45, No. 5 (1948), pp. 443–59.

"In the comparisons of both the *total* population and the *active* population the Kinsey figures are higher in nearly all cases. Though the specific differences are not statistically significant the trend of difference is consistent in direction."* (Keep footnote as is.)

Pages 112–113. In the third line on p. 113 change the phrase "would have been much greater" to "would probably have been greater."

Page 138–139. This long quotation from my review would have been more fair had I added the paragraph from KPM (pp. 415–416) that is quoted on p. 141 of CMT

* * *

PAUL WALLIN, *Stanford University*[1]

THE position taken by CMT in regard to most of my criticisms of KPM is entirely acceptable to me. There is little, if any, disagreement between us on the issues discussed in my review. Indeed I feel impelled to register dissent with only one of their comments on my work.

This concerns my criticism of KPM for reporting hundreds of means, and medians based on less than 50 cases, many on less than 10 and some on less than 4. CMT observe (pp. 88–89) that

> Perhaps the criticism is not totally justified. There seems to be no general reason for throwing away a sample of size 20 or 30 just because it is small, especially when it may be the only data on the subject In many fields of endeavor a sample of 30 is regarded as quite substantial (testing A-bombs for example). It does seem reasonable, however, to choose a minimum sample size for a cell, and report smaller numbers only with some danger signal attached to warn the reader of the instability of the entry.

In view of the difficulty of securing subjects for the KPM kind of research I would agree that samples of size 20 or 30 should not necessarily be discarded because of their smallness. However, given the great range of variability of human behavior—particularly in as large and complex a society as the United States—I question the value of using non-random samples of this magnitude—even if accompanied by a danger signal—as a basis for deriving national statistics on various sexual practices of American males. I likewise am dubious about the accuracy of means or medians calculated from 10 or fewer cases as estimates of the frequency of different types of sexual behavior within what CMT call the "sampled population" as distinguished from the "target population."

[1] "An Appraisal of Some Methodological Aspects of the Kinsey Report," *American Sociological Review*, Vol. 15, No. 2 (1949), pp. 197–210.

In testing A-bombs considerable confidence might be attached to results achieved even with a sample of 1, since bombs are constructed under controlled conditions intended to maximize the probability of homogeneity in their behavior. Unfortunately for the problems it poses for social scientists, the behavior of human beings entails many degrees of freedom and is thus likely to be much more heterogeneous than that of A-bombs.

Other than the above, there are two very minor points which I should like to comment on briefly.

(1) In discussing the KPM comparison of the original and retake histories of 162 subjects I said that inconsistency between the two sets of histories "would suggest that the data being sought cannot be accurately reported by the subjects The evidence of witnesses whose testimony varies from one time to another is of questionable accuracy." CMT note (p. 96) that if the latter statement

> . . is taken literally it implies that all readings of voltmeters, balances and other physical instruments are of "questionable accuracy." If so, this would not be a serious criticism

It hardly need be said that the differences alluded to in my statement were *significant* differences rather than differences of *any* degree. I pointed out in a subsequent paragraph that " the data indicate test-retest reliability is high on incidence items but far below acceptable standards on frequency items "

(2) In discussing the possible bias in the KPM sample due to refusals, I stated "That the number of non-volunteers was considerable is indicated by the fact that 'Perhaps 50,000 persons have heard about the research through lectures and perhaps half of the histories now in hand [i.e 6000 out of 12,000] have come in consequence of such contacts' (KPM, p. 38) The authors' comment on another study is applicable to their own. They say it is " . . . open to the very severe criticism that it involved only a highly selected sample of the total population. What is more serious, one is left guessing as to the histories of the 51 per cent that failed to answer the questionnaire' (KPM, p. 619)."

CMT assumed I was implying here that KPM had a refusal rate of 88% (44,000 non-volunteers out of 50,000 asked to volunteer for interviews). Consequently, it is CMT's calculation, and not my more general inference that " . . . the number of non-volunteers was considerable . . . ," which KPM have characterized (p. 61) as

> . . . meaningless, since many of the 50,000 were not approached in any way to give histories.

* * *

W. ALLEN WALLIS, *University of Chicago*[1]

THE CMT report is excellent. In thoroughness, competence, objectivity, wisdom, originality, and lucidity it leaves nothing to be desired It is a credit not only to its authors but to the Association that sponsored it and the officers who selected the authors.

The purpose of this comment, however, is not to review CMT in general, but to consider any disagreements between CMT and my own review of KPM, which they scrutinize word by word—some words several times, in fact. The cynic will sense from my opening paragraph that in the CMT report I have found not substantial criticism of my review but repeated agreement and approval Nevertheless, by checking every listing of my name in the index, I have found half a dozen direct or indirect demurrers to points of mine They, and my comments on them, are all so small and dull that the reader would do well to skip directly from here to (or perhaps beyond) my last paragraph.

(1) CMT point out (p. 86) that it would be difficult to make meaningful tests of the significance of such differences as those between the KPM interviewers. They are right, though it is an exaggeration to say that every interviewer would have had to visit every collecting site While my remark does refer to the first Kinsey report, it occurs, as the first sentence they quote shows, in connection with suggestions for future work Furthermore, it is not as clear to me as it seems to be to CMT that KPM in pages 133–43 are not attempting to assess the statistical significance of the differences whose smallness they emphasize. The CMT position could be supported if one were to give KPM credit for reasonably precise and careful use of statistical terms; for example, they call the differences "not material," but do not call them "not significant" They do, however, compare differences in means with the variability of individual observations, they do refer to "the independent interviewing of three persons drawing their samples very largely at random . . ," they do imply a sign test between two interviewers, and they do draw at least one clear conclusion from their sample of three interviewers to "any other group of interviewers" (all on p 135). Altogether, in this and similar comparisons, they create in my mind some atmosphere of significance-testing and some impression that differences, by and large, either are not significant or can be explained away by other variables

(2) CMT are right in believing (p. 107) that I make a distinction in meaning between "fail to provide confidence" and "destroy confidence" or "contradict."

[1] "Statistics of the Kinsey report," *Journal of the American Statistical Association*, 44 (1949), pp 463–84

(3) I cannot agree with CMT's contention (p. 109) that KPM are "correct, informative and helpful" when they say that "a few high-rating individuals affect the means more than [many] low-rating individuals"; I would have to know how high-rating are the few individuals and how numerous are the low-rating individuals in the comparison. It is likely, of course, that KPM were groping for the correct idea which CMT state, that in distributions skewed to the right, the mean exceeds the median, but my criticism pertained to the confusion of their presentation.

(4) I agree with CMT (p. 110) that necessary and sufficient conditions for equality of the mean and median need not be discussed by KPM. My illustration was simply one of several to indicate the presence in the KPM report of many a "matter like a misspelled word, which an author would change if his attention were called to it." To see how the author might want to change this one, I suggested two possible interpretations that occurred to me, but indicated dissatisfaction with both.

(5) My advice to make comparisons among the component groups of a whole rather than between a component and the whole related to planning methodological checks in future work. One of the comparisons to which I alluded, that between age groups, may have been inappropriate, since this comparison was not made for methodological purposes (though the advice may be appropriate to it); and this ill-chosen allusion seems to have diverted CMT's attention from the context of the passage they quote (p. 111). The remarks they make—they avoid taking a position—are, however, valid if taken out of context.

(6) That CMT analyze so carefully (pp. 111–12) my references to sequential procedures suggests that they credited me with having something deeper in mind than I did. In referring to the problem of determining sample size, I stated that it may be impractical in the case of averages to specify the required accuracy as a fixed percentage of the true figure. The point here is that frequently it is technically impossible to specify a fixed sample size that, with prescribed confidence, meets an accuracy requirement so specified. This kind of technical difficulty is sometimes surmountable by sequential methods. Stein's double-sampling procedure for which CMT give the bibliographical details is the one I assumed my mention of double-sampling would bring to the minds of statisticians; his sequential procedure, however, was then very new and virtually unknown, so I cited it more specifically. CMT point out that Stein's sequential procedure was not available when KPM's work was done; I mentioned it in discussing the

determination of sample sizes for future work (pp. 467–8 of my review).

These six points are trivial, but they are the only ones I can find where CMT seem to take even this much issue with the points made in my review. As far as I can see, the differences between us are negligible.

Turning to a broader issue, CMT apply one criterion to KPM's work that merits further consideration. They compare KPM's statistical methods with those of other studies of sex, rather than other studies involving surveys. Other studies of sex are not numerous and, judging by the descriptions by KPM (pp 23–34) and by William O Jenkins (CMT, Appendix B), not good More important, those few that are in any way comparable with KPM's were published at least a decade before KPM's, and the intervening decade was one of extraordinarily rapid progress in techniques of sampling, and of analyzing samples from, human populations Thus, the comparison with other sex studies has a little the character of proclaiming a champion on the basis of knocking down a straw man On the other hand, it is unrealistic to expect workers in any field to keep abreast of the latest methodological developments, however vital to their own research. Furthermore, Kinsey's interviewing started a decade before the KPM report was published, and many of the methods for whose absence the KPM report is criticized had not been developed nearly so far then as now. Perhaps the best statistical methods used previously in the particular field of application provide an appropriate lower bound on what it is reasonable to expect, and the best methods used previously in any field an appropriate upper bound.

* * *

WILLIAM G COCHRAN, *The Johns Hopkins University*
FREDERICK MOSTELLER, *Harvard University*
JOHN W TUKEY, *Princeton University*

WE APPRECIATE the opportunity given us by the Editor of the *Journal* to add to this discussion But we feel that our own views have already been stated at considerable length. We wish to thank the reviewers for their thoughtful discussions and useful amplifications. Although we are not in complete agreement with all the points made, our remaining differences hinge mostly on questions of judgment, emphasis, or frame of reference.

THE EFFICIENCY OF DOUBLE SAMPLING FOR ATTRIBUTES

H C. HAMAKER AND R. VAN STRIK
Philips Research Laboratories, N V. Philips' Gloeilampenfabrieken, Eindhoven, Netherlands

§1 INTRODUCTION

IN A random walk diagram where we plot the total number inspected as abscissa and the number of defectives observed as ordinate, a double sampling plan is represented by two screens S_1 and S_2 as indicated in Fig. 1 When after inspection of a first sample of n_1 items the random walk ends in section A_1 of screen S_1 the lot is accepted, and

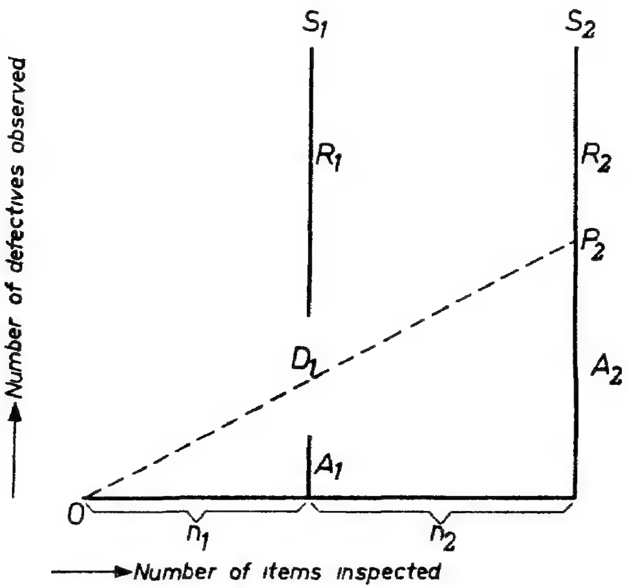


FIG. 1. Random walk diagram for a double sampling plan.

when it ends in R_1 the lot is rejected But whenever we arrive somewhere in the opening D_1 the case is dubious; we proceed to take a second sample of size n_2 and pass a final judgment according to the sections A_2 and R_2 of screen S_2. Evidently a double sampling plan requires 5 parameters for its full specification; two sample sizes and three "decision numbers."

If we wish to replace a single sampling plan by a double sampling plan we shall as a rule require both plans to possess nearly the same operating characteristics (OC). Since a single sampling plan is fixed by two parameters and the double plan by 5 there will be a considerable variety of double sampling plans satisfying this requirement, and the choice of the most suitable one among these becomes a complex problem.

In a way this problem was solved by the Statistical Research Group [6] and in the subsequent Military Standard 105A [4] and since the

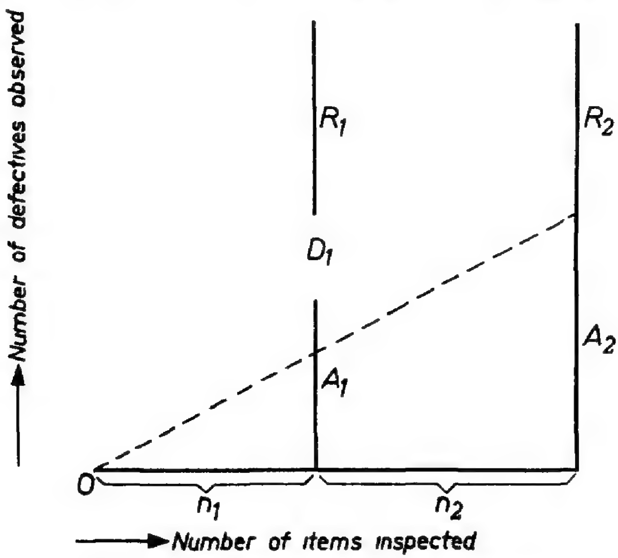


FIG. 2. Random walk diagram for an intuitively inefficient double sampling plan.

tables contained in this standard have been widely accepted we may conclude that the solution is a satisfactory one

Yet it may be doubted whether it is in every respect the best solution conceivable, as the following argument will show. Apart from random deviations a random walk created by the inspection of items taken from a homogeneous lot will move along a straight line through the origin. Hence if we draw a straight line from the origin to the dividing point P_2 in the second screen (Fig. 1) this line should preferably pass somewhere through the center of the opening D_1 For with a set of screens as depicted in Fig. 2 the judgments based on the first and

the total sample are not properly balanced. We accept some lots by the first sample which would have a high chance of being rejected should we proceed to take a second sample; and we will likewise be led to consider some lots as doubtful by the first sample which are very likely to be rejected after inspection of a second sample. Intuitively we conclude that double plans with a set up as in Fig. 2 must be inefficient.

Generally we must therefore require that in a double sampling plan

$$(c_1 + \tfrac{1}{2}) < \frac{n_1}{n_1 + n_2}(c_3 + \tfrac{1}{2}), \tag{1a}$$

$$(c_2 + \tfrac{1}{2}) > \frac{n_1}{n_1 + n_2}(c_3 + \tfrac{1}{2}), \tag{1b}$$

where n_1 and n_2 are the sample sizes and c_1, c_2, and c_3 the three decision numbers. Throughout we shall use the lower decision numbers, so defined that a lot is accepted when the number of defectives is c_1 or less in the first, or c_3 or less in the total sample; and a second sample is taken when in the first sample we have observed a number of defectives greater than c_1 but *less than or equal* to c_2.

Since we accept at c_3 and reject at c_3+1 defectives we have assumed the dividing point in the screen S_2 to be at $(c_3+\frac{1}{2})$, etc

In the Mil. Std. 105A $c_2=c_3$ and condition (1b) is consequently always satisfied. Since in these tables $n_2=2n_1$ condition (1a) becomes

$$(c_1 + \tfrac{1}{2}) < (c_3 + \tfrac{1}{2})/3 \tag{2}$$

and many of the double sampling plans incorporated in the Mil. Std 105A do not satisfy this last criterion. This suggests that it is perhaps possible to find other double sampling plans with nearly the same O.C. but a better efficiency.

The original aim of the researches described in this paper was to determine how far this intuitive argument is correct and whether the improvements, if any, would be worth practical consideration. In the course of our work it appeared, however, that there are many other aspects of double sampling, such as the use of truncation in the second sample and the choice of parameters which have to be taken into account to complete the picture. Hence all these factors will be discussed in due course.

In §§2, 3, and 4 we begin by developing a theory of double sampling efficiency which already has been described briefly in earlier papers [2, 3] and will form the basis of all our later arguments.

12. "EQUIVALENT" SAMPLING PLANS

It will be convenient to use the following short-hand notation for double sampling plans

$$D(n_2/n_1;\, c_1,\, c_2,\, c_3), \tag{3}$$

where n_2/n_1 denotes the ratio of the sample sizes, and c_1, c_2, and c_3 the three decision criteria For example the symbol

$$D(2;\, 2,\, 6,\, 10) \tag{4}$$

designates all double sampling plans where the second sample is twice as large as the first; and by which a lot is accepted, when the first

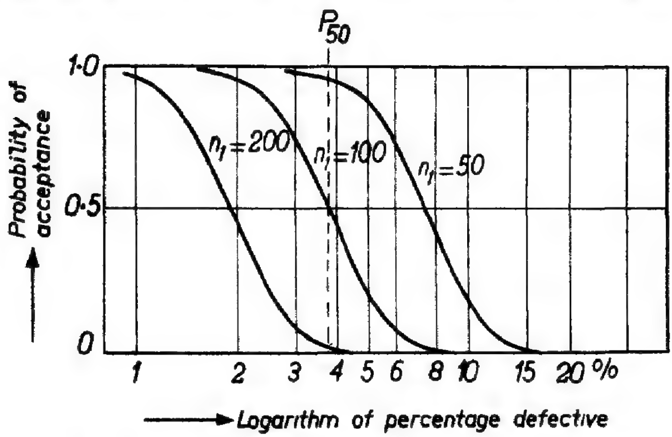


FIG. 3. The operating characteristic plotted against log p. A change of sample size produces a horizontal shift but no other changes.

sample contains *2 or less* defectives, or the total sample *10 or less* defectives, while the lot is rejected when the first sample contains *more than 6*, or the total sample *more than 10* defectives

By the symbol (3) or (4) a double sampling plan is fixed apart from the size n_1 of the first sample As long as we assume Poisson conditions—which hold true in most industrial applications—the probability of acceptance for a plan described by (3), will be a function of n_1p, where p is the fraction defective in the lot. The most convenient way to use this fact is by plotting the OC, not against p as is usually done, but against log p as done in Fig. 3. A change in n_1 will then correspond to a shift of the characteristic parallel to itself. Hence, if we wish to compare different sampling plans, we can always start by adjusting the

sample sizes so that the operating characteristics have a specific point in common, for example the *point of control*, or indifference quality, p_{50} defined by

$$P(p_{50}) = 0.5, \tag{5}$$

where P is the probability of acceptance.

If we wish to investigate the efficiency of double as compared to single sampling, the first step is to find two plans which possess OC's as nearly alike as possible; for otherwise the inspection performances achieved by the two plans are different and the numbers of observations required are not comparable. To this end we start by changing the sample sizes so that the two operating characteristics possess the same point of control. In Fig. 4, for example, the OC for the plan

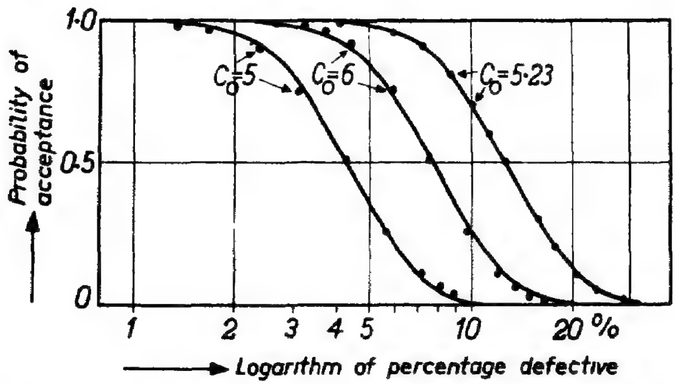


FIG. 4. The OC for the double sampling plan $D(2; 5, 10, 10)$ compared, after an appropriate adjustment of the sample size, with those of the single sampling plans for $c_0=5$, 6, and 5.23.

D (2; 5, 10, 10) is compared with those for single plans with $c_0=5$ and $c_0=6$, after this adjustment has been performed; both give a reasonably close fit and at first sight it is difficult to judge which of the two is best.

In order to achieve some precision in our argument it is therefore necessary to effect the adjustment by a more accurate mathematical treatment. To this end we have adopted the principle that two operating characteristics may be considered as *equivalent* when they not only possess the same point of control p_{50}, but also the same slope at that point. It will be convenient to use the slope of the curve when p is plotted on a logarithmic scale, because this slope is, under Poisson

conditions, independent of the absolute value of the sample sizes. Hence we have defined as a second parameter the *relative slope*

$$h = -\left(\frac{d \ln P}{d \ln p}\right)_{p_{50}} = -2\left(p \frac{dP}{dp}\right)_{p_{50}}, \tag{6}$$

where P is the probability of acceptance, and we will call two OC's "equivalent" when they possess the same values of h and p_{50}.[1]

Even so a complete adjustment cannot be attained, because the acceptance criteria, and h in consequence, vary in discrete steps. But this difficulty can be overcome by introducing broken values of the acceptance number c_0 of the single sampling plan. If, while drawing a sample, we determine at the same time by a lottery with probabilities 0.77 for a 5 and 0.23 for a 6, whether the acceptance number shall be 5 or 6, such a plan may be defined as belonging to an acceptance number 5 23, and its operating characteristic is obtained by linear interpolation between the curves for $c_0 = 5$ and $c_0 = 6$ Such a method is not recommended for actual practice, but in a theoretical argument it enables us to find a single sampling plan with exactly the same slope h as a given double sampling plan, and consequently with a very close approach in its operating characteristic, as Fig. 4 shows. We might similarly introduce broken values of the sample size n_0 in order to achieve perfect adjustment in the point of control p_{50} Usually, however, sample sizes are so large that the error committed in rounding n_0 to an integer value may be neglected The introduction of broken values of c_0 is more essential.

The adjustment is greatly facilitated by the fact that the parameters p_{50} and h are, for single sampling plans, connected to the sample size n_0 and the acceptance number c_0 by two very simple relations [2, 3],

$$n_0 p_{50} = c_0 + 0.67, \tag{7}$$

$$\frac{\pi}{2} h^2 = n_0 p_{50} + 0.06 = c_0 + 0.73. \tag{8}$$

The method of comparison is now as follows For a given double sampling plan we find the values of h and p_{50} by numerical computation, and then obtain n_0 and c_0 from (7) and (8); these data specify an *equivalent* single sampling plan Details of the numerical computation are given in an Appendix.

---

[1] In earlier papers [2, 3] the notation p_0, h_0 has been used instead of p_{50}, h The advantage of using p_{50} is obvious, with p_{90} or p_{10} we can then describe other points on the OC It is not necessary to indicate the slope at the same time as h_{50} because we will consider the slope only at the point p_{50}.

Equivalence, as here defined, does not mean, of course, that the OC's are completely coincident; near the top or the bottom of the curves slight differences remain But equivalence does mean that the OC's are on the whole so close to each other that the differences have no appreciable influence in practical applications; for practical purposes the OC's may be considered as coincident.

13. AN ALTERNATIVE DEFINITION OF EQUIVALENCE

Commonly OC's are not specified by the parameters p_{50} and h, but by two fractions defective, say, p_{95} and p_{10} corresponding to acceptance probabilities of 0.95 and 0.10 respectively.

For single sampling plans and under Poisson conditions the ratio p_{10}/p_{95} is, like the slope h, a function of the acceptance number c_0 alone. Hence an "equivalent" single sampling plan might alternatively

TABLE 1

COMPARISON OF TWO DEFINITIONS OF "EQUIVALENCE"

Double Sampling Plan	Single Sampling Plan Adjusted by Means of				n_0/n_0'
	p_{50}, h		p_{10}, p_{95}		
	n_0	c_0	n_0'	c_0'	
D (2; 0, 1, 1), $n_1=90$	139	0.63	132	0.69	1.053
D (2, 0, 4, 4), $n_1=90$	238	3.66	231	3.64	1.030
D (2; 5, 13, 13), $n_1=150$	196	6.96	203	7.64	0.966
D (2, 2, 9, 9), $n_1=90$	202	7.23	194	7.18	1.041
D (2; 1, 4, 8), $n_1=75$	192	6.88	185	6.69	1.038

be defined as the plan with the same value of p_{10}/p_{95} as the double plan considered and with a sample size n_0 so adjusted that the two OC's coincide at p_{95} and p_{10}

For a set of 5 different double sampling plans the two definitions of equivalence considered above are compared in Table 1. The last set of three are double sampling plans with approximately the same OC's but with different decision numbers. The plan $D(2; 5; 13, 13)$ corresponds to a random walk diagram as in Fig 2, that is, to a presumably inefficient choice of the acceptance numbers; the plan $D(2, 2, 9, 9)$ still adheres to the $c_2=c_3$ principle but the acceptance numbers have been better chosen, while for the last plan in Table 1, $c_2 \neq c_3$

The differences between the two definitions of equivalence appear to be slight; they amount to not more than 5 per cent of the sample size.

The OC's of the plans occurring in one row in Table 1 lie very close together, so close in fact that it would be difficult to draw them clearly separate in one figure. For the sampling plan D (2; 2, 9, 9) the differences are shown in Table 2, for the other plans they were of the same order of magnitude. As might be expected the plan adjusted by means of p_{50} and h lies closer to the double sampling plan in the middle range of P values but further away at large or at small values of P.

TABLE 2

ACCEPTANCE PROBABILITIES FOR D (2; 2, 9, 9) AND THE EQUIVALENT SINGLE SAMPLING PLANS ADJUSTED BY p_{50}, h, AND BY p_{10}, p_{95} RESPECTIVELY

Percentage Defective	Plan		
	D (2; 2, 9, 9) $n_1=90$	$c_0=7.23$ $n_0=202$	$c_0=7.18$ $n_0=194$
	P		
1%	0 9996	0.9992	0 9993
2	.9636	.9555	9625
3	.7688	.7623	7903
4	.4734	4733	5155
5	2431	2340	.2711
6	1159	.0963	.1196
7	0551	0343	0458
8	.0266	.0109	0157

On the whole both definitions of equivalence seem equally acceptable. In the following we have used the definition of §2 because it is easier in numerical computations and has the advantage of formula's (7) and (8). If another definition is adopted this may cause systematic changes of 3 to 5 per cent in the efficiencies as defined in the next section. Such changes will not alter the main conclusions at which we shall arrive.

§4. A THEORY OF EFFICIENCY

By means of p_{50} and h we can now find an equivalent single sampling plan in a reasonably precise manner and we proceed to compare the average number of observations required by equivalent sampling plans.

The ratio of the average sample size of a double sampling plan to the sample size of its equivalent single sampling plan will be called *inverse efficiency*. The inverse efficiency measures the amount of inspection under double sampling in terms of the amount of inspection required by a single sampling plan with the same inspection performance In keeping with general practice we assume that truncation may be permitted in the second sample of a double sampling plan but is not allowed in the first sample or in single sampling. In this section we investigate the case where no truncation is used; the effect of truncation is discussed in §§5 and 6.

The average sample size of a double sampling plan, and the inverse efficiency in consequence, vary with the fraction defective in the lot inspected Actually we get more convenient curves if we plot the inverse efficiency not against p but against the corresponding probability of acceptance P, using a normal probability scale The curves thus obtained will be termed *efficiency characteristics;* these we shall use to study the efficiency of various double sampling plans.

The double plan $D(2; 5, 10, 10)$ does not satisfy the criterion (2) of §1 and must by the arguments of that section intuitively be classified as inefficient. This view is strikingly corroborated by its efficiency characteristic as shown in Fig 5. Even for good lots with an acceptance probability above 0 99 the inverse efficiency is hardly less than unity and the average amount of inspection is for the double sampling plan almost the same as for the equivalent single sampling plan. For lots of poor quality the double sampling plan is much worse.

If we now change over from the plan $D(2, 5, 10, 10)$ to the plan $D(2, 2, 10, 10)$ we must by the argument of §1 expect an improvement in efficiency and this too is brought out by Fig 5.

By the efficiency characteristics in Fig 5, two double sampling plans are compared each with its own equivalent single sampling plan. Can these curves also be interpreted as portraying the efficiency of the two double sampling plans with respect to one another? We believe they can for the following reason: The two double plans under discussion possess different slopes of their operating characteristics, and consequently the average sample number curves for these plans are not comparable Generally, however, we know that the steeper an OC the larger the number of observations required, and this holds true for single, double, and sequential sampling alike. Can we measure the influence of the slope and correct for it?

Numerical computation yields for the plan $D(2; 2, 10, 10)$ a relative slope $h=2.427$, and for the plan $D(2; 5, 10, 10)$ $h=1.949$. Hence by

equation (8) the ratio of the equivalent single sample sizes turns out to be

$$\frac{n_0}{n_0'} = 1.56.$$

To achieve the greater slope of the OC of the plan $D(2;\ 2,\ 10,\ 10)$ would require under single sampling 1 56 times as many observations. If we assume that this factor holds for double sampling as well, we are in a position to eliminate the difference in slope and thereby to render

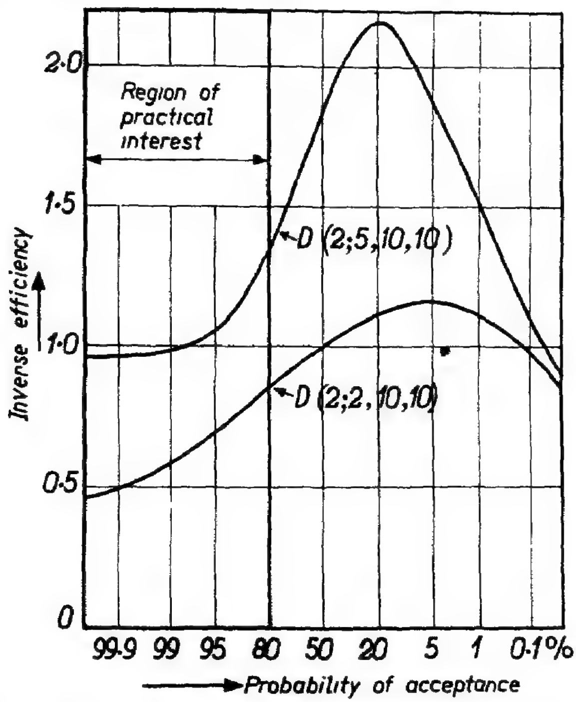


FIG 5. Efficiency characteristics for the double plans $D(2;\ 5,\ 10,\ 10)$ and $D(2;\ 2,\ 10,\ 10)$. By reducing the acceptance number c_1 from 5 to 2 a striking improvement is achieved

average sample number curves for the two double plans more directly comparable. This is exactly what is achieved by the inverse efficiency as introduced above

We may put the same argument in a somewhat different way. Single sampling plans with different acceptance numbers have different

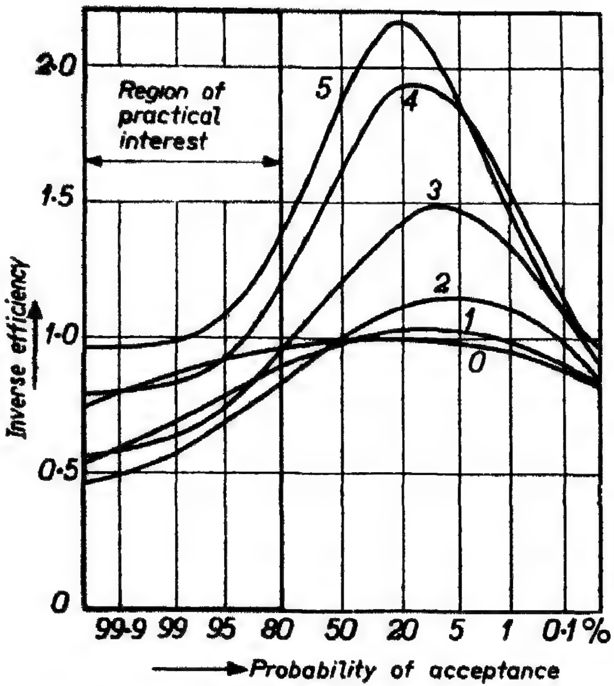


Fig. 6. Efficiency characteristics for the set of plans $D(2; c_1, 10, 10)$ when $c_1 = 5, 4, \cdots, 0$

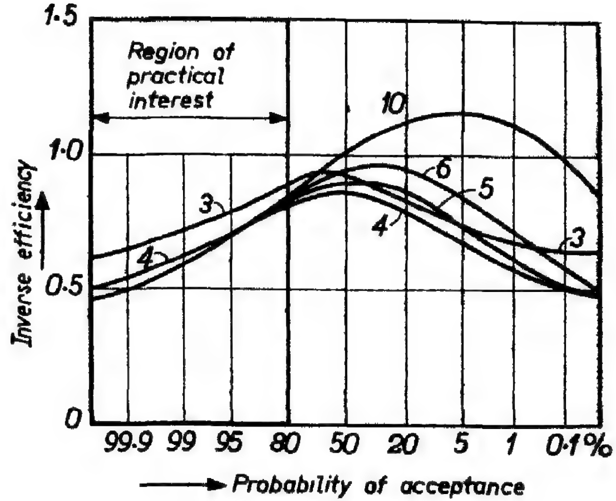


Fig. 7. Efficiency characteristics for the plans $D(2; 2, c_3, 10)$ with $c_3 = 10, 6, 5, 4$, and 3.

relative slopes, and from the point of view of efficiency they can not be compared. It is therefore not unreasonable to accept single sampling plans as a general standard of reference which has by definition unit efficiency; we can then refer all other sampling plans to this common standard.

From this point of view Fig. 5 can be accepted as giving a reasonable picture of the relative merits of the two double sampling plans with respect to one another, and efficiency characteristics can be used to judge the merits of any double sampling plan.

Fig. 5, however, may be misleading in another respect. In practical situations the majority of the lots submitted for inspection is accepted and only a small fraction of the lots is rejected. This means that most of the lots have a high probability of acceptance, say, 95 per cent or thereabout at least. Hence, in practice we will operate mainly on the left-hand side of Fig. 5 and the striking peaks of the efficiency curves are in reality of little interest because lots falling in that region are rare. Thus, without express warning Fig. 5 may easily lead to a biased interpretation. To avoid this, the part of Fig. 5 with probabilities of acceptance greater than 0.80 has been specially indicated as the *region of practical interest* In comparing different sampling plans we should pay attention to this part of the figure mainly. As a further application of efficiency characteristics we consider in Fig. 6 the set of plans $D(2; 5, 10, 10)$, $D(2; 4, 10, 10)$, $D(2; 3, 10, 10)$, etc. We see at once that in the region of practical interest the plan $D(2; 2, 10, 10)$ is the best.

Next, fixing c_1 at 2, we study the further improvement that can be obtained if we abandon the $c_2 = c_3$ principle (Fig. 7) As expected a change from $D(2; 2, 10, 10)$ to $D(2; 2, 5, 10)$ gives an improved efficiency mainly on the side of poor lots; in the region of practical interest the effect is almost nil and both plans are equally satisfactory. If we reduce c_2 still further the opening in the first screen of the random walk diagram becomes too narrow and the efficiency is impaired.

§5 TRUNCATION

The efficiency characteristics so far considered were based on the assumption that all the samples are completely inspected To reduce the number of observations truncation of the second sample is sometimes recommended; that is, inspection of the second sample is stopped as soon as the total number of defectives observed exceeds the final acceptance number c_3. Evidently truncation will not alter the OC but will improve the efficiency. This is again nicely demonstrated by the efficiency characteristics.

To study the influence of truncation we have used the three double sampling plans already considered in Table 1, viz., $D(2;\, 5,\, 13,\, 13)$, $D(2;\, 2,\, 9,\, 9)$ and $D(2;\, 1,\, 4,\, 8)$. These have nearly the same relative slopes, $h=2.21$, 2.25 and 2.20 respectively, and if the sample sizes are properly adjusted, as in Table 1, they have nearly identical OC's.

The random walk diagrams are given in Fig. 8 The plan $D(2;\, 5,\, 13,\, 13)$ is of the type of Fig 2, and is consequently not really efficient. With a plan of this type lots accepted as doubtful by the first sample have a high chance of being rejected after the second sample, and we may consequently expect truncation to produce a considerable improvement With the plan $D(2;\, 2,\, 9,\, 9)$ the improvement will presumably be less, though still pronounced because the upper half of the opening in the first screen is still too large. Finally, with the plan $D(2;\, 1,\, 4,\, 8)$ we should expect that truncation will have only a small influence

The actual efficiency characteristics in Fig. 9 are in complete agreement with these predictions The efficiencies for the truncated plans were computed from the formulas given in "*Sampling Inspection*" [6, p. 209].

It will be noted that in the region of practical interest the effect of truncation is relatively unimportant. In fact, truncation is only permitted in the second sample and a second sample occurs frequently only with relatively poor lots. In actual practice second samples are rare and consequently the reduction in the amount of inspection resulting from truncation is small.

Whether we use truncation or not the double plan $D(2;\, 2,\, 9,\, 9)$ is always considerably better than the plan $D(2;\, 5,\, 13,\, 13)$; the latter forms part of Mil Std. 105A and Fig 9 illustrates the improvement that can be achieved by a better choice of acceptance numbers.

§6. TRUNCATION AND THE CHOICE OF DOUBLE SAMPLING PLANS

Dodge and Romig's original sampling tables [1] contain double sampling plans for which $c_2=c_3$ and since then all double sampling plans proposed for practical use have been of the same type. Dodge and Romig did not mention that a specific choice was involved and so far little or no attention has been paid to this problem. It has been discussed in a report by Stein and Shaw [7] which is, however, not readily accessible and is not cited in the main textbooks on the subject. Hence we shall devote some space to it here.

The choice $c_2=c_3$ can be justified as simplifying the specification of a double sampling plan, while the loss in efficiency occurs in a region

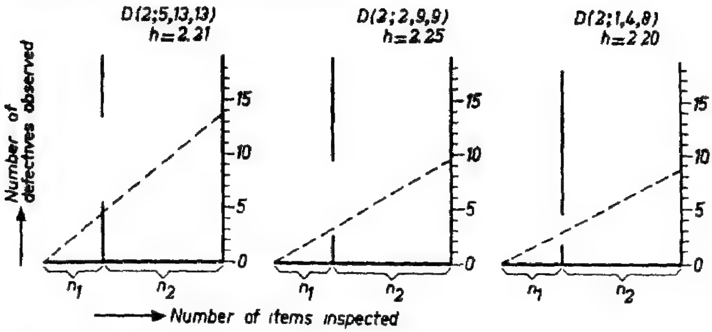


FIG 8 Random walk diagram for three double plans with nearly the same slope h The efficiency of these plans without and with truncation is shown in Fig 9

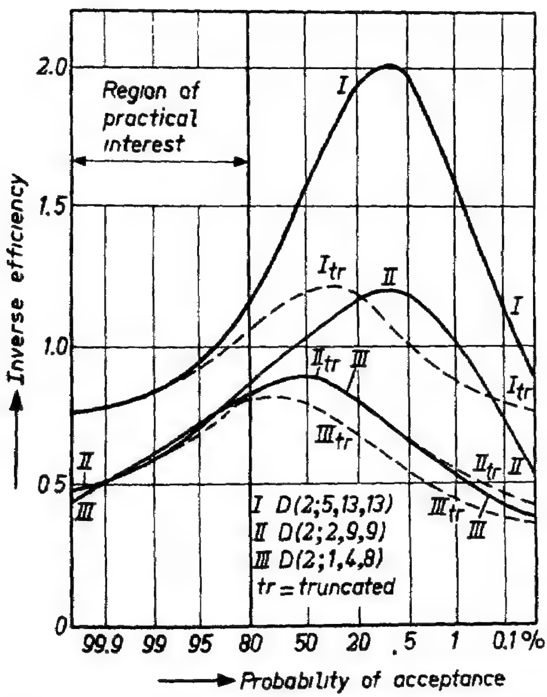


FIG 9. The effect of truncation on efficiency. Efficiency characteristics for the plan $D(2; 5, 13, 13)$, $D(2; 2, 9, 9)$ and $D(2, 1, 4, 8)$ both without and with truncation.

of small practical importance (Fig. 9). On the other hand if we take $c_2=c_3$ and then recommend truncation we simplify on one hand and complicate matters on the other, and it may well be questioned whether this is correct. In fact we see from Fig 9 that the plan $D(2; 2, 9, 9)$ *with truncation* has nearly the same efficiency characteristic as $D(2; 1, 4, 8)$ *without truncation*. Which of these two plans is to be preferred in practice?

In the Mil. Std. 105A the acceptance and rejection numbers are given in full; the entries corresponding to the two sampling plans just mentioned would read for $D(2; 2, 9, 9)$

	Accept	*Reject*
First sample	2	10
Second sample	9	10

and for $D(2; 1, 4, 8)$

	Accept	*Reject*
First sample	1	5
Second sample	8	9

the advantage of $c_2=c_3$ being that the rejection numbers are the same in both samples We do not believe, however, that this is a real practical advantage. As a rule the operator inspects a sample setting defectives apart; when inspection is completed he counts the number of defectives and then consults the table as to the decision to be taken. He will probably have to consult the table after the second as well as after the first sample, and if so there is no specific advantage in having the same rejection numbers in both cases.

With truncation it is another matter. Truncation is only permitted in second samples and these occur infrequently. The routine will be to complete the inspection of a sample and count the number of defectives afterwards But with truncation the inspector will suddenly have to keep an eye on the number of defectives whenever he has to inspect a second sample. Hence truncation necessitates irregular and infrequent breaks in the normal inspection routine, which will be inconvenient in practice.

Hence we are of the opinion that if the acceptance and rejection numbers are given as fully as in the Mil. Std. 105A sampling plans with $c_2\neq c_3$ will be more satisfactory in practice than plans with $c_2=c_3$ combined with truncation.

It may be added that we do not know how far truncation is really applied. It is recommended in most textbooks, but in Mil. Std. 105A, which is primarily meant for immediate practical use, truncation is not mentioned at all. We are inclined to conclude from this that trunca-

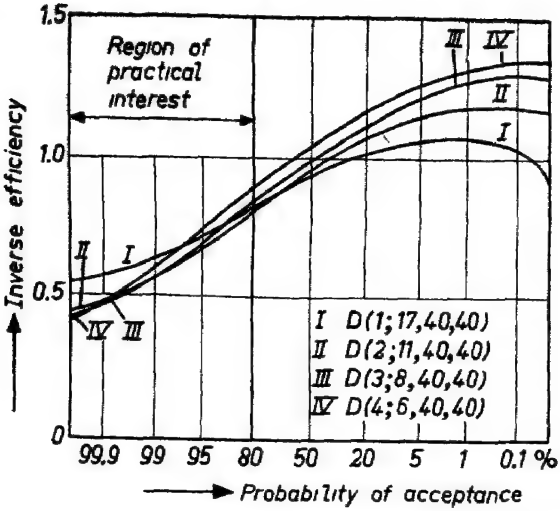


FIG. 10. The effect of the ratio of the sample sizes, when $c_2 = c_3$.

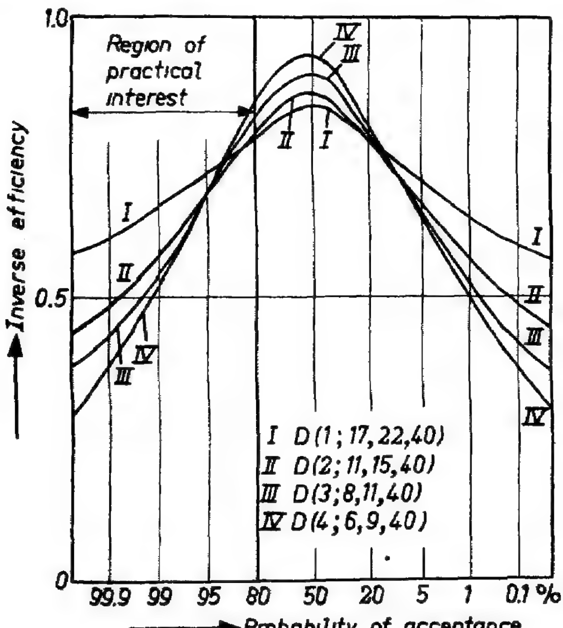


FIG. 11. The effect of the ratio of the sample sizes when $c_2 \neq c_3$.

tion has been found inconvenient; perhaps for the reasons we have given above. Some practical information on this point would be desirable.

In conclusion it may be emphasized once more that in the region of practical interest the gain from either truncation or $c_2 \neq c_3$ is small anyhow.

§7 THE RATIO OF THE SAMPLE SIZES

In the original double sampling tables of Dodge and Romig [1] the ratio of the sample sizes, n_2/n_1, varied somewhat but nearly all later tables have used a constant ratio, $n_2/n_1=2$ By the method of the foregoing sections we can now easily settle in how far this is a satisfactory choice

In Figs. 10 and 11 we give the efficiency characteristics for sampling plans with $n_2/n_1=1$, 2, 3, and 4. In computing these curves large acceptance numbers have been used so that c_1 and c_2 can be fairly accurately adjusted to give optimum efficiency We see that for good lots the greater the ratio n_2/n_1 the better the efficiency At a probability of acceptance of 95 per cent the value of n_2/n_1 has comparatively little influence and to the right of this point a high ratio n_2/n_1 leads to decreased efficiency. The differences in this region are larger for $c_2=c_3$ than for $c_2 \neq c_3$ plans (bearing in mind a difference in the scales of Figs 10 and 11). All this is reasonable enough.

On the whole $n_2/n_1=2$ seems a satisfactory choice When quality is very good and the rejection of lots rare, a higher ratio may be of some advantage, but it may then be still better to stop sampling inspection altogether.

We have now discussed the efficiency of double sampling plans in all in various aspects In future papers we intend to deal with the choice of a sampling plan in practice and to provide a set of tables which enable a rapid choice of appropriate single and double sampling plans

APPENDIX

FORMULAS FOR THE COMPUTATION OF p_{50} AND h FOR DOUBLE SAMPLING PLANS

We shall write

$$q(k;m) = \frac{e^{-m}m^k}{k!}, \tag{9}$$

$$R(c;m) = \sum_{0}^{c} q(k,m), \tag{10a}$$

$$Q(c,m) = 1 - R(c;m) = \sum_{k=c+1}^{\infty} q(k;m) \tag{10b}$$

We then have

$$\dot{q}(k; m) = \frac{dq(k; m)}{dm} = q(k - 1; m) - q(k; m), \tag{11}$$

$$\dot{Q}(c, m) = \frac{dQ(c; m)}{dm} = + q(c, m) = Q(c - 1; m) - Q(c; m), \tag{12}$$

formulas which can be used also for $k=0$ or $c=0$ if we define

$$q(-1, m) = 0, \qquad Q(-1, m) = 1 \tag{13}$$

The probability of acceptance for a double sampling plan with $c_2 = c_3$ is

$$P = R(c_1, m_1) + \sum_{k=c_1+1}^{c_2} q(k, m_1)R(c_2 - k; m_2), \tag{14}$$

where

$$m_1 = n_1 p, \qquad m_2 = n_2 p,$$

and p = the fraction defective in the lot

Furthermore we have

$$R\{c_2, (m_1 + m_2)\} = \sum_{k=0}^{c_2} q(k, m_1)R(c_2 - k, m_2) \tag{15}$$

and making use of this relation (14) can be rewritten

$$P = R(c_1, m_1) + R(c_2, m_1 + m_2) - \sum_{k=0}^{c_1} q(k; m_1)R(c_2 - k, m_2), \tag{16}$$

or introducing Q instead of R

$$P = 1 - Q(c_2, m_1 + m_2) + \sum_{k=0}^{c_1} q(k; m_1)Q(c_2 - k, m_2) \tag{17}$$

Since for a specific plan the sample sizes n_1 and n_2 are fixed we have for varying p

$$h = - 2p\frac{dP}{dp} = - 2m_1\frac{dP}{dm_1}, \tag{18}$$

and

$$\frac{dm_2}{dm_1} = \frac{n_2}{n_1}. \tag{19}$$

Differentiation of (17) yields

$$\begin{aligned} - \dot{P} = - \frac{dP}{dm_1} &= q(c_1, m_1)Q(c_2 - c_1 - 1; m_2) \\ &+ \left(1 + \frac{n_2}{n_1}\right)\left[q(c_2; m_1 + m_2) + \sum_{k=0}^{c_1} q(k, m_1)Q(c_2 - k, m_2)\right. \\ &\left.- \sum_{k=0}^{c_1} q(k; m_1)Q(c_2 - k - 1, m_2)\right] \end{aligned} \tag{20}$$

Since the functions $q(k, m)$ and $Q(c, m)$ have been tabulated in extenso by Molina [5], formulas (17) and (20) are convenient for practical use. The computation of a sum of products from 0 to c_1 involves as a rule less labor than from (c_1+1) to c_2, particularly when $c_2 = c_3$.

When $c_2 \neq c_3$ we have

$$P = R(c_1; m_1) + \sum_{k=c_1+1}^{c_2} q(k; m_1)R(c_3 - k; m_2)$$

or

$$P = 1 - Q(c_2; m_1) - \sum_{k=c_1+1}^{c_2} q(k; m_1)Q(c_3 - k, m_2), \tag{21}$$

and by differentiation

$$
\begin{aligned}
-\dot{P} = -\frac{dP}{dm_1} &= q(c_2; m_1) - q(c_2; m_1)Q(c_3 - c_2 - 1; m_2) \\
&+ q(c_1; m_1)Q(c_3 - c_1 - 1; m_2) \\
&+ \left(1 + \frac{n_2}{n_1}\right)\left[\sum_{k=c_1+1}^{c_2} q(k; m_1)Q(c_3 - k - 1; m_2)\right. \\
&\left. - \sum_{k=c_1+1}^{c_2} q(k, m_1)Q(c_3 - k; m_2)\right]
\end{aligned}
\tag{22}
$$

It is now no longer possible to convert the summations but this is not necessary either, because when $c_2 \neq c_1$, c_1 and c_2 are lying fairly close together and a summation of products from c_1+1 to c_2 is brief and simple.

The general policy was to compute P and $-\dot{P}$ for two values of m_1 which gave P values in the immediate neighborhood of $P=0.50$, and then to find the correct values by inverse interpolation. Table 3 gives full details of the numerical procedures.

TABLE 3

ILLUSTRATING THE NUMERICAL COMPUTATION OF n_1p_{50} AND h FOR THE SAMPLING PLAN $D(2, 1, 4, 8)$

m_1 m_2	2.9 5.8	3.0 6.0
$a=1-Q(c_2; m_1)$	0.831777	0 815263
$b=q(c_2; m_1)$	0.162154	0.168031
$c=q(c_2, m_1)Q(c_3-c_2-1; m_2)$	0.134594	0.142624
$d=q(c_1, m_1)Q(c_3-c_1-1; m_2)$	0 057701	0 058803
$e=\sum_{k=c_1+1}^{c_2} q(k; m_1)Q(c_3-k; m_2)$	0.311792	0.332528
$f=\sum_{k=c_1+1}^{c_2} q(k, m_1)Q(c_3-k-1; m_2)$	0 409015	0.426992
$P=a-e$	0 519985	0.482735
$-\dot{P}=b-c+d+3(f-e)=$	0.376930	0.367602

By inverse interpolation

$m_{11}=n_1p_{50}$	2.954
$-\dot{P}_{50}$	0.3726
$h=-2\,m_{11}\dot{P}_{50}$	2.201
c_0	6.88
n_0p_{50}	7.551
n_1/n_0	0.392

The value of $m_{11}=n_1p_{50}$ was obtained by inverse linear interpolation: the OC in the region between $P=0.52$ and 0.48 is sufficiently straight to arrive at an accurate value that way.

To obtain the derivative $\dot{P}_{10}$ a somewhat more accurate method of interpolation was found necessary For two values of m_1, $m_{10}=2.9$ and $m_{12}=3.0$, we have computed the corresponding values P_0, P_2, and $\dot{P}_0$, $\dot{P}_2$. We may then adjust a third degree equation so that it passes through the points m_{10}, P_0, and m_{12}, P_2 and has the correct slopes. From this equation we can next derive the slope at the point m_{11}, $P_1=\frac{1}{2}$. This principle yields the interpolation formula

$$\dot{P} = \left(\frac{P_2 - P_0}{m_{12} - m_{10}}\right) 6\Theta\Phi + \dot{P}_0\Theta(1 - 3\Phi) + \dot{P}_2\Phi(1 - 3\Theta),$$

where

$$\Theta = \frac{m_{11} - m_{10}}{m_{12} - m_{10}}, \qquad \Phi = (1 - \Theta)$$

This formula we have used. Originally a cruder method of computation was adopted but this led to slight discrepancies which could only be removed by using a more accurate technique By using two copies of Molina's tables so that the tables of individual and cumulative Poisson distributions can be laid side by side, the entire computation can be carried out straight from the tables Routine computations of h for a particular double sampling plan do not take more than half an hour. It should be noted that

$$P_M(c, m) = Q(c - 1, m) \tag{23}$$

when P_M denotes the cumulative probabilities as tabulated by Molina

At any value of m_1 the probability of a second sample is

$$\sum_{k=c_1+1}^{c_2} q(k, m_1) = Q(c_1 + 1, m_1) - Q(c_2; m_1); \tag{24}$$

hence the inverse efficiency is

$$\text{I E} = \frac{n_1}{n_0}\left[1 + \frac{n_2}{n_1}\left\{Q(c_1 + 1; m_1) - Q(c_2, m_1)\right\}\right] \tag{25}$$

while the corresponding probability of acceptance is provided by (17) or (21).

REFERENCES

[1] Dodge, Harold F , and Romig, Harry G., *Sampling Inspection Tables*, New York; John Wiley and Sons, Inc. 1944: or *Bell System Technical Journal* 8 (1928) 613–31.

[2] Hamaker, H. C., "Some notes on lot-by-lot inspection by attributes," *Review of the International Statistical Institute* 18 (1950), 179–96.

[3] Hamaker, H. C., "The theory of sampling inspection plans," *Philips Technical Review*, 11 (1950), 260–70.

[4] *Military Standard 105A. Sampling Procedures and Tables for Inspection by Attributes.* U. S. Government Printing Office The Standard is reproduced in full in "Quality Control and Industrial Statistics" by A. J. Duncan, Chicago, R. D. Irwin Inc , 1952.

[5] Molina, Edward C., *Poisson's Exponential Binomial Limit*, New York; Van Nostrand, 1947.

[6] Statistical Research Group, Columbia University, *Sampling Inspection*, New York; McGraw-Hill Book Co , 1948.

[7] Stein, Arthur, and Shaw, Lawrence W., "Some methods of reducing the amount of inspection in the application of double sampling inspection procedures." *Ballistic Research Laboratory Report No. 248*, Aberdeen Proving Ground, 1943.

GENERALITY OF CONFIDENCE INTERVALS FOR A REGRESSION FUNCTION

EDWIN L CROW*

U. S. Naval Ordnance Test Station, China Lake

In obtaining a confidence interval for the ordinate to a regression line (or surface, more generally) it is commonly assumed in statistics books, either explicitly or implicitly in the course of derivation, that the "independent" variable X is assigned fixed values, i e., that in the hypothetical repeated sampling encompassed by the probability model the values of X must be the same as those in the sample actually observed (e.g , Cramér [3, pp 548–54]. On the other hand it is fairly common to find applications of such confidence intervals in which the values of X evidently arose at random and would not be repeated in a further sample, except under hypothetical restriction, usually this restriction is not mentioned Hald [6, pp 522, 627] gives a unified treatment in which the independent variable is explicitly allowed to take either fixed or random values, but he does not demonstrate why results obtained in the case of fixed values (pp 528–40) also apply in the case of random values (p 616) It seems desirable here explicitly to bridge this gap, and it is easily done The underlying idea is contained in Bartlett's [2] consideration of conditional variation of a sample and in Fisher's [5] consideration of samples having the same *configuration* (cf [8]).

The argument used is familiar in the theory of confidence intervals Suppose, for example, that a random sample $(x_1, y_1, \cdots, x_n, y_n)$ is drawn from a bivariate normal population of (X, Y), and we calculate from the theory appropriate for fixed $(x_1, \cdots, x_n)$ a confidence interval with confidence coefficient γ for the ordinate to the regression line at a prescribed value of X. It follows that in indefinitely repeated hypothetical sampling, in which each sample is taken from the normal population $(X_1, Y_1; \cdots; X_n, Y_n)$ subject to the condition $(X_1, \cdots, X_n) = (x_1, \cdots, x_n)$, the proportion of confidence intervals covering the true regression ordinate is γ But this is true whatever the x_i; hence the probability that the confidence interval covers the true regression ordinate (without restriction on the x_i) is

$$\int \cdots \int \gamma f(x_1, \cdots, x_n) dx_1 \cdots dx_n = \gamma,$$

* Now at the Boulder Laboratories of the National Bureau of Standards.

where $f(x_1, \cdots, x_n)$ represents the joint probability density function of the n random sample values of X In other words, the confidence interval calculated for fixed values of X is a confidence interval with the same confidence coefficient without restriction on the values of X. The proof is seen to consist merely of averaging a constant, γ, over the distribution of all possible samples of X of size n.

The same reasoning can be stated more generally, so that it applies to a multiple regression ordinate, to a total or partial regression coefficient [1, 3, pp. 550, 553], and to the difference between two regression coefficients for two samples from possibly different multivariate normal populations With respect to the last result it is interesting to note that Bartlett [1] specifically retained the requirement of fixed values of X in the two samples, although he obtained the more general result for regression coefficients in a single sample In general (as suggested by Edward A. Fay), for any distribution $F(x, y)$ of (generally vector-valued) random variables X and Y, if $R(Y|x)$ is a confidence region for a parameter θ having confidence coefficient γ with respect to each conditional distribution of Y given that $X=x$, then $R(Y|X)$ (where X is a random variable) is a confidence region for θ having the same confidence coefficient with respect to the joint distribution F

The analogous remarks can be made for tolerance [11] and prediction [9] intervals.

It would appear that the introduction of additional random fluctuation, even though it be only through the values of X, should be reflected somehow in the inference, whereas the above intervals are formally identical Thus the variance of the regression coefficient in bivariate normal regression is $n/(n-3)$ times the corresponding variance with x values held fixed (as may be seen from [3, pp 402, 549], where the sample variance $\sum(x_i-\bar{x})^2/n$ in the latter case is also equal to the population variance), but when the regression coefficient is "Studentized" to obtain a t statistic, the same result is obtained for the two cases. With the same confidence coefficient the confidence intervals are identical in the two cases, but the probabilities of covering some value in the parameter space other than the true value are different. Likewise, with the same significance level (probability of a Type I error) the tests of a hypothetical value of a regression coefficient which can be derived from the corresponding confidence intervals are formally identical, but the probabilities of a Type II error differ

The preceding statement can be illustrated by reference to the operating characteristic of the t-test graphed for a significance level of 0 05 by Ferris, Grubbs, and Weaver [4]. Suppose a sample of size n' is drawn

from a bivariate normal population (X, Y) for which the regression coefficient of Y on X is β, the standard deviation of x is σ_x and the standard deviation of the Y deviations from the regression line of Y on X is σ. It is desired to test the null hypothesis H_0: $\beta=o$, while actually $|\beta|=\lambda'\sigma$ The conditional probability of accepting H_0 for any particular fixed set of x's can be found from the referenced graph by taking $n=n'-1$, $\lambda=\lambda's_x$, where $s_x^2=\sum(x_i-\bar{x})^2/n$ For example, if $n'=5$, $\lambda'=1$ 5, $s_x=1$, then the graph yields the probability 0.47. The unconditional probability of a Type II error can be obtained by numerically integrating the product of such a probability (read from the graph) and the probability density of $s_x^2=\sigma_x^2\chi_n^2/n$ where χ_n^2 is chi-square with $n=n'-1$ degrees of freedom. The integral is a function of σ_x; if $\sigma_x=1$ in the above example, the result by Simpson's Rule is 0.54, appreciably larger than the conditional probability 0 47 obtained with $s_x=1$. If the median value 0 916 of s_x is used for comparison, the conditional probability is 0 53, but the 5 and 95 per cent points of s_x yield conditional probabilities of 0.15 and 0 84. Thus the probability of a Type II error in testing a regression coefficient in the bivariate normal case can hardly be calculated using the sample s_x if the sample size is small. Patnaik [10] has given four numerical examples of the entire operating characteristics (actually the complementary power function curves) for both the fixed and random cases in the analogous problem of one-way analysis of variance In every example the operating characteristic for the fixed case lies below the other, and Patnaik conjectured that this would be true in general. Johnson [7] confirmed Patnaik's conjecture for the usual range of significance levels (0.001 to 0.10) but showed it unlikely to be universally true

Welch [12] has pointed out, with the aid of an example, that a test based on the distribution in samples of the same configuration, as above, may not be the best possible test, even though it is the best possible test within any one configuration.

REFERENCES

[1] Bartlett, M. S, "On the theory of statistical regression," *Proceedings of the Royal Society of Edinburgh*, 53 (1933), 260–83.

[2] Bartlett, M S, "Properties of sufficiency and statistical tests," *Proceedings of the Royal Society*, A, 160 (1937), 268–82.

[3] Cramér, Harald, *Mathematical Methods of Statistics* Princeton: Princeton University Press, 1946.

[4] Ferris, Charles D., Grubbs, Frank E, and Weaver, Chalmers L., "Operating characteristics for the common statistical tests of significance," *Annals of Mathematical Statistics*, 17 (1946), 178–97, in particular Fig. 6, p. 195

[5] Fisher, R. A., "Two new properties of mathematical likelihood," *Proceedings of the Royal Society*, A, 144 (1934), 285–307.

[6] Hald, A., *Statistical Theory with Engineering Applications*. New York: John Wiley & Sons, 1952.

[7] Johnson, N. L., "Comparison of analysis of variance power functions in the parametric and random models," *Biometrika*, 39 (1952), 427–29.

[8] Kendall, Maurice G., *The Advanced Theory of Statistics*, Vol II, London: Charles Griffin and Company Limited, 1946. Secs 21 47, 22, 19–21.

[9] Mood, Alexander McFarlane, *Introduction to the Theory of Statistics*. New York, Toronto, London. McGraw-Hill Book Company, Inc., 1950. Pp. 297–9, 304–5.

[10] Patnaik, P. B., "The non-central χ^2- and F-distributions and their applications," *Biometrika*, 36 (1949), 202–32, 226–27 in particular).

[11] Wallis, W. Allen, "Tolerance intervals for linear regression," in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, Berkeley and Los Angeles University of California Press, 1951 Pp 43–51.

[12] Welch, B. L., "On confidence limits and sufficiency, with particular reference to parameters of location," *Annals of Mathematical Statistics*, 10 (1939), 58–69.

DISTRIBUTIONS OF SOLUTIONS OF A SET OF LINEAR EQUATIONS (WITH AN APPLICATION TO LINEAR PROGRAMMING)*

M. M Babbar
University of Costa Rica

This article presents an approach to deriving distributions of the variables representing the solution of a set of simultaneous linear equations, when the coefficients are subject to random errors In addition, the distribution of a linear function of these variables is also discussed Specifically, the model considered is as follows:

$$(B+b)X = (Q+\epsilon),$$

where B is an $m\times m$ nonsingular matrix of known constants and b is an $m\times m$ matrix of random errors, Q is an m-column vector of known constants and ϵ is an m-column vector of the corresponding random errors The problem is to derive the distribution of a general element of X and also the distribution of a linear function.

$$(C+c)'X,$$

where C is an m-column vector of known constants and c an m-column vector of the corresponding random errors

A similar problem of errors in solutions of a set of linear equations when the coefficients are subject to error has been considered by many authors, for example, Lonseth [14], Hotelling [12], and Turing [18], but most of them have been concerned with "rounding" errors Dwyer has an elaborate section in his book *Linear Computations* [7] devoted to the problem of the solution of linear equations with coefficients subject to errors, but the errors again are not considered to be random and the treatment, therefore, is not based on the theory of probability Recently, Box and Hunter [2], by an extension of the argument given by Fieller [8], have presented a formula for obtaining a confidence region for the solution of a set of linear equations where the errors in the coefficients have a multivariate normal distribution The general

* This article is based on the author's dissertation [1] submitted to the Graduate Faculty of Iowa State College in partial fulfillment of the requirements for the degree of Doctor of Philosophy He is indebted to Gerhard Tintner for suggesting the problem and for guiding in its solution and to Earl O Heady for setting up the empirical example and helping in its analysis Also, he acknowledges the improvements due to the suggestions and criticism made by other members of the staff of the departments of economics, statistics, and mathematics of Iowa State College

derivation does not assume the errors to be uncorrelated, and relatively easier formulas are derived when the errors are assumed to be uncorrelated. If the equations are poorly "conditioned," however, there may be serious difficulties in the empirical application of the results.

In the following pages, the treatment of the problem is rather different. Actual approximate distributions of the variables which occur as a solution of the set of equations as well as the distribution of a linear function are derived. Also, direct and easily applicable formulas are presented to find the confidence limits of the variables and the linear function Moreover, the probability that any variable or the linear function assumes a value greater than or less than any preassigned number can be investigated easily.

The approach outlined in this paper, however, involves a high degree of approximation, though situations are mentioned in which it will be exact. The compensating aspect is that the analysis can be applied without much tedious computation or tabulation.

Problems involving the solution of a set of linear equations occur frequently in many applied fields In economics, for example, Leontief-type [13] input-output analysis depends on such a solution In the general area of linear programming, Dorfman [6], making certain assumptions which guarantee non-degeneracy, has shown that the solution which maximizes the linear objective function will involve activities, active or disposal, equal in number to the linear restrictions. Thus, if the variables irrelevant to the planned program are removed, the solution can be written as

$$BX = Q.$$

When this plan is put in practice, however, the "operating" values of the input coefficients represented by the elements of B may be different from their "anticipated" values used in planning. Thus, if random fluctuations in these coefficients are allowed, the operating model will be of the form

$$(B + b)X = (Q + \epsilon).$$

Therefore the analysis of this paper can be usefully employed to predict the variation in the values of X and also to predict the variation in the linear objective function In Section III a detailed example is discussed to illustrate this application

The treatment in Sections I and II is, however, quite general and it could be applied to experimental designs and problems of the physical sciences which involve the solution of linear equations

I. REDUCTION OF THE MODEL

A. *Consider a set of m linear equations in m variables.*

$$(B + b)X = (Q + \epsilon) \tag{1.1}$$

where B is an $m \times m$ non-singular matrix whose elements are known constants and Q is an m-column vector whose elements (q_i) are also known. Small b represents an $m \times m$ matrix of random errors whose elements (b_{ij}) have known probability distributions, such that

$$\left.\begin{array}{l} E(b_{ij}) = 0 \\ \text{and} \\ E(b_{ij}^2) = \sigma_{ij}^2 \end{array}\right. \quad \text{for} \begin{cases} i = 1, 2, \cdots, m \\ \\ j = 1, 2, \cdots, m \end{cases}.$$

E denotes the mathematical expectation.

Matrix $(B+b)$ is also assumed to be non-singular.

Since the elements of the matrix b are, later on, assumed to be normally distributed on the real line, this assumption that $(B+b)$ be non-singular can be achieved with probability one.

Further, ϵ is an m-column vector of errors (ϵ_i) whose elements have known probability distributions such that

$$\left.\begin{array}{l} E(\epsilon_i) = 0 \\ \text{and} \\ E(\epsilon_i^2) = \tau_i^2. \end{array}\right. \quad \text{for} \quad i = 1, 2, \cdots, m$$

Our aim is to derive approximate distributions for the variables $x_1, x_2, \cdots, x_m$ which occur in the model as a solution of the linear equations (1.1).

Let

- $|B|$ denote the determinant of the matrix (B_{ij}),
- $|B+b|$ denote the determinant of the matrix $(B_{ij}+b_{ij})$,
- β_{ij} denote the co-factor of the element B_{ij} in $|B|$,
- $|D^k|$ denote the determinant of matrix (B_{ij}) when its k-th column is replaced by the column (q_i),
- D_{ij}^k denote the co-factor of the element in the i-th row and j-th column of $|D^k|$,
- $|D^k+d^k|$ denote the determinant of the matrix $(B_{ij}+b_{ij})$ when its k-th column is replaced by column vector $(q_i+\epsilon_i)$. We know that the solution of equations (1.1) is given by

$$x_k = \frac{|D^k + d^k|}{|B + b|} \qquad \text{for } k = 1, 2, \cdots, m. \tag{1.2}$$

Also, we will be interested in finding the distribution of a linear function of the solutions x_i, $i=1, 2, \cdots, m$. For this purpose, let C be an m-column vector of fixed constants $(C_1, C_2, \cdots, C_m)$ and c be a column vector of errors $(c_1, c_2, \cdots, c_m)$, where $E(c_i)=0$, $i=1, 2, \cdots, m$, and $E(c_i^2)=\omega_i^2$, $i=1, 2, \cdots, m$. We consider the linear function

$$
\begin{aligned}
y &= \sum_{r=1}^{m} (C_r + c_r)x_r \\
&= \frac{1}{|B+b|}\sum_{r=1}^{m} (C_r + c_r)(|D^r + d^r|).
\end{aligned}
\tag{1.3}
$$

B. *Rule of procedure for an approximation.* Now the determinants involved in expressions (1.2) and (1.3) will be reduced to approximate expressions using the following rule of procedure.

We will ignore all cross products of errors of the second and higher orders in the expansion of a determinant whose elements involve errors.

It may be noticed that although our rule of procedure does not assume the square and higher powers of the errors to be negligible, they do not occur in the expansion of the determinants anyway, since no element is multiplied by itself in the expansion of a determinant.

C. *Approximate formulas.* Using the rule of procedure given above, we get expansions of the determinants involved in expressions (1.2) and (1.3) as follows:

$$
|D^k + d^k| \cong |D^k| + \sum_{i=1}^{m} \beta_{ik}\epsilon_i + \sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq k}}^{m} D_{ij}^k b_{ij} = N(x_k) \text{ say.} \tag{1.4}
$$

Similarly,

$$
|B+b| \cong |B| + \sum_{i=1}^{m}\sum_{j=1}^{m} \beta_{ij}b_{ij} = D(x) \text{ say.} \tag{1.5}
$$

Now,

$$
E(|D^k + d^k|) \cong |D^k| = \delta_k \text{ say,} \tag{1.6}
$$

$$
V(|D^k + d^k|) \cong \sum_{i=1}^{m} (\beta_{ik})^2\tau_i^2 + \sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq k}}^{m} (D_{ij}^k)^2\sigma_{ij}^2 = \sigma_k^2. \tag{1.7}
$$

V denotes the variance.

Again,

$$
E(|B+b|) \cong |B| = \beta \text{ say,} \tag{1.8}
$$

and

$$(1.9) \qquad V(|B+b|) \cong \sum_{i=1}^{m}\sum_{j=1}^{m}(\beta_{ij})^2\sigma_{ij}^2 = \sigma_B^2 \text{ say.}$$

Also

$$(1.10) \qquad \theta(|D^k+d^k|, |B+b|) \cong \sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq k}}^{m} D_{ij}^k\beta_{ij}\sigma_{ij}^2 = \sigma_{Bk} \text{ say.}$$

θ denotes the covariance.

We are interested in finding the distribution of the quotient corresponding to (1.2)

$$(1.11) \qquad x_k = \frac{N(x_k)}{D(x)}\cdot$$

For the reduction of expression (1.3) we have,

$$(1.12) \qquad \sum_{r=1}^{m}(C_r+c_r)(|D^r+d^r|) \cong \sum_{r=1}^{m}C_r|D^r| + \sum_{r=1}^{m}C_r\left(\sum_{i=1}^{m}\beta_{ir}\epsilon_i\right) + \sum_{r=1}^{m}C_r\sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq r}}^{m}D_{ij}^r b_{ij} + \sum_{r=1}^{m}|D^r|c_r = N(y) \text{ say,}$$

$$(1.13) \qquad E(N(y)) = \sum_{r=1}^{m}C_r|D^r| = \delta_y \text{ say}$$

$$(1.14) \qquad V(N(y)) \cong \sum_{r=1}^{m}\left(\sum_{i=1}^{m}C_i\beta_{ri}\right)^2\tau_r^2 + \sum_{r=1}^{m}C_r^2\sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq r}}^{m}(D_{ij}^r)^2\sigma_{ij}^2 + \sum_{r=1}^{m}|D^r|^2\omega_r^2 = \sigma_{N(y)}^2 \text{ say}$$

Similarly $|B+b| = D(x) = D(y)$ say (Ref. 1 5) where $E(D(y)) = \beta$ and $V(D(y)) = \sigma_B^2$ and

$$(1.15) \qquad \theta[N(y), D(y)] \cong \sum_{r=1}^{m}C_r\sum_{i=1}^{m}\sum_{\substack{j=1\\ j\neq r}}^{m}D_{ij}^r\beta_{ij}\sigma_{ij}^2 = \sigma_{B\cdot N(y)} \text{ say,}$$

and the problem again is to find the distribution of the quotient.

$$(1.16) \qquad y = \frac{N(y)}{D(y)}.$$

It is important to notice the situations in which the above formulas will be exact.

First, if $b \equiv 0$ and $c \equiv 0$ and the elements of ϵ are uncorrelated, all the formulas given above will be exact.

Second, if $c \equiv 0$ and if all the elements of b except those in one column, say the k^{th} column, are identically equal to zero and the interest is only x_k, then the formulas relevant to x_k will be exact.

In these two cases, if the errors are not uncorrelated but their covariances are known, the formulas for the variances and covariances of the numerator and the denominator can be generalized to include the relevant covariance terms In that case again we get an exact analysis, for the two cases mentioned above.

II. DISTRIBUTIONS

Suppose that the probability distributions of the elements of b, ϵ and c are known In Section I, it has been seen that we want the distribution of a quotient of two linear functions of the errors involved To find this, we could find the distributions of the numerator and denominator separately and then find the distribution of the quotient In general, this could be done by writing down the joint distribution of the errors involved and making the necessary transformations Alternatively, the same could be accomplished by employing characteristic functions and the inversion formulas presented by Gurland [11]. Another way of doing the same thing may be by evaluating the moments of the distribution as suggested by K Pearson [15] and C. C. Craig [4] All of these procedures, however, seem difficult, particularly when the number of error terms involved is large and the distributions of the errors are non-normal. If the errors are normally distributed then one can deal with the problem of finding the distributions of expressions (1 11) and (1 16) as follows.

A *Probability Distributions.* It may be noticed that in both the expressions the numerators $N(x_k)$ and $N(y)$ and the denominators $D(x)$ and $D(y)$, are linear functions of the normally distributed errors Such functions are themselves normally distributed. $N(x_k)$ will be normally distributed with expected value δ_k and variance σ_k^2 $N(y)$ will be normally distributed with expected value δ_y and variance $\sigma_{N(y)}^2$ Also $D(x)$ and $D(y)$ will be normally distributed with expected value β and variance σ_B^2 In this way, the problem reduces itself to finding the dis-

tribution of a quotient of two normal variates. We will use the following theorem proved by R. C. Geary [10].

Theorem If N and D are normally distributed variables with $E(N)$ and $E(D)$ as the mean values, σ_N^2, σ_D^2 as variances and σ_{ND} as covariance, and $Z=N/D$ is the quotient, then the expression,

$$t = \frac{E(D)Z - E(N)}{(\sigma_N^2 - 2\sigma_{ND}Z + \sigma_D^2 Z^2)^{1/2}} \tag{2.1}$$

is approximately normally distributed with mean zero and variance one, provided

$$E(D) > 3\sigma_D, \tag{2.2}$$

i.e., the coefficient of variation of D is $<1/3$.

In proving this theorem, Geary shows that if (2.2) is satisfied then the probability that Z lies between any two real numbers Z_1 and Z_2 is practically the same as the probability that a standard normal variate lies between t_1 and t_2, the values corresponding to Z_1 and Z_2 obtained from (2 1). In the particular case when $E(D)=3\sigma_D$ the difference between the two probabilities for all values of Z_1 and Z_2 will be less than .0027. If $E(D)$ is greater than $3\sigma_D$, this difference will be less than a number smaller than .0027. Assuming such difference to be negligible, Geary concludes the truth of the theorem The importance of the condition (2.2) for the efficiency of the approximation is, therefore, apparent.

Now, using this theorem we get the probability distribution of Z as follows:

$$\begin{aligned} f(Z)dZ &= \frac{1}{\sqrt{2\pi}} \frac{[E(D)\sigma_N^2 - E(N)\sigma_{ND}] + Z[E(N)\sigma_D^2 - E(D)\sigma_{ND}]}{[\sigma_N^2 - 2\sigma_{ND}Z + \sigma_D^2 Z^2]^{3/2}} \\ &\quad \times e^{-1/2[\{ZE(D)-E(N)\}^2/\sigma_N^2 - 2\sigma_{ND}Z + \sigma_D^2 Z^2]} \times dZ. \end{aligned} \tag{2.3}$$

Identifying Z with x_k we get the probability distribution of x_k

$$\begin{aligned} f(x_k)dx_k &= \frac{1}{\sqrt{2\pi}} \frac{[\beta\sigma_k^2 - \partial_k\sigma_{Bk}] + x_k[\partial_k\sigma_B^2 - \beta\sigma_{Bk}]}{[\sigma_k^2 - 2\sigma_{Bk}x_k + \sigma_B^2 x_k^2]^{3/2}} \\ &\quad \times e^{-1/2[(x_k\beta - \partial_k)^2/\sigma_k^2 - 2\sigma_{Bk}x_k + \sigma_B^2 x_k^2]} \times dx_k. \end{aligned} \tag{2.4}$$

Again identifying z with y we get the corresponding probability distribution:

$$(2.5)\quad f(y)dy = \frac{1}{\sqrt{2\pi}} \frac{[\beta\sigma_{N(y)}{}^2 - \partial_y\sigma_{BN(y)}] + y[\partial_y\sigma_B{}^2 - \beta\sigma_{BN(y)}]}{[\sigma_{N(y)}{}^2 - 2\sigma_{BN(y)}y + \sigma_B{}^2y^2]^{3/2}}$$
$$\times\, e^{-1/2[(y\beta-\partial_y)^2/\sigma_{N(y)}{}^2 - 2\sigma_{BN(y)}y + \sigma_B{}^2y^2]} \times dy.$$

B. *Confidence limits.* In practice, we may not be interested in the theoretical distributions as suggested in Section II A. In this section, we consider probability limits of the quotient $Z=N/D$ obtained directly without writing down the distributions. We have seen that

$$t = \frac{E(D)Z - E(N)}{(\sigma_N{}^2 - 2\sigma_{ND}Z - \sigma_D{}^2Z^2)^{1/2}}$$

is approximately normally distributed with mean zero and variance unity. If we want limits with 100α per cent probability, where the probability coefficient α is a fraction, then consulting the standard normal tables, we find a positive number γ such that the probability of the absolute value of the standard normal variate being greater than γ is $(1-\alpha)$. In other words

$$(2.6)\qquad P\left[\left|\frac{E(D)Z - E(N)}{(\sigma_N{}^2 - 2\sigma_{ND}Z - \sigma_D{}^2Z^2)^{1/2}}\right| \leqq \gamma\right] = \alpha$$

or

$$P[\{E(D)Z - E(N)\}^2 \leqq \gamma^2\{\sigma_N{}^2 - 2\sigma_{ND} + \sigma_D{}^2Z^2\}] = \alpha$$

or

$$P[\{E(D)^2 - \gamma^2\sigma_D{}^2\}Z^2 - 2\{E(D)\cdot E(N) - \gamma^2\sigma_{ND}\}Z + \{E(N)^2 - \gamma^2\sigma_N{}^2\} \leqq 0] = \alpha$$

P stands for a probability statement
Thus the roots of the quadratic equation in Z:

$$(2.7)\qquad \{E(D)^2 - \gamma^2\sigma_D{}^2\}Z^2 - 2\{E(D)\cdot E(N) - \gamma^2\sigma_{ND}\}Z - \{E(N)^2 - \gamma^2\sigma_N{}^2\} = 0$$

will give the two numbers and Z lies between those limits with 100α per cent probability.

Applying this general formula to find the probability limits of x_k and y of Section I, we see that the roots of the quadratic equation in Z:

$$(2.8)\qquad \{\beta^2 - \gamma^2\sigma_B{}^2\}Z^2 - 2\{\partial_k\beta - \gamma^2\sigma_{Bk}\}Z + \{\partial_k{}^2 - \gamma^2\sigma_k{}^2\} = 0$$

will give the probability limits for solution x_k (for $k=1, 2, \cdots, m$) and the roots of equation

$$\{\beta^2 - \gamma^2\sigma_B{}^2\}Z^2 - 2\{\partial_y\beta - \gamma^2\sigma_{BN(y)}\}Z + \{\partial^2 y - \gamma^2\sigma_{N(y)}{}^2\} = 0 \tag{2.9}$$

will give the probability limits for the linear function y In both cases, the probability is 100α per cent.

C. *Cumulative distributions.* If the frequency distribution of a continuous variable is known, its cumulative distribution is obtained by integration from the lower limit of the range of variation to another variable For instance, in Section II A it was indicated how to write down the distributions of x_k and y, that is, the distributions of the elements of X individually as well as that of a linear function of these elements The variable $Z=N/D$ where N and D are normal variates, can represent both functions If $f(Z)$ represents the frequency function of Z, then $F(v)$, the corresponding cumulative distribution function, is obtained as follows.

$$F(v) = \int_{-\infty}^{v} f(Z)dZ \tag{2.10}$$

In practice, however, it may be too tedious to evaluate the integral. Therefore, we suggest the following procedure, based on a paper by Fieller [9] He has shown that the chance of obtaining a value of the variable $Z=N/D$ not less than v, that is $\{1-F(v)\}$, can be computed as below:

$$\{1 - F(v)\} = \int_h^\infty \int_k^\infty N(\rho)dxdy + \int_{-h}^\infty \int_{-k}^\infty N(\rho)dxdy \tag{2.11}$$

where $N(\rho)$ is bivariate (x and y) normal distribution with means zeros, both variances equal to unity and correlation coefficient ρ. The constants h, k and ρ are computed as follows·

$$h = \frac{E(D)}{\sigma_D} \tag{2.12}$$

$$k = \frac{E(N) - vE(D)}{[\sigma_N{}^2 - 2\sigma_{ND}v + \sigma_D{}^2v^2]^{1/2}} \tag{2.13}$$

and

$$\rho = \frac{\sigma_{ND} - v\sigma_D{}^2}{\sigma_D(\sigma_N{}^2 - 2\sigma_{ND}v + \sigma_D{}^2v^2)^{1/2}}. \tag{2.14}$$

The values of the integrals in (2.11) can be obtained from K. Pearson's tables [16]. The tables are available only for positive values of h and k. However, the following relations regarding the bivariate normal distribution with the means zero and the variances unity and correlation coefficient ρ, can be used in case either or both of h and k are negative.

$$\int_{-h}^{\infty}\int_{k}^{\infty} N(\rho)dxdy = \int_{k}^{\infty} N(0,\,1)dy - \int_{h}^{\infty}\int_{k}^{\infty} N(-\rho)dxdy \tag{2.15}$$

$$\int_{h}^{\infty}\int_{-k}^{\infty} N(\rho)dxdy = \int_{h}^{\infty} N(0,\,1)dx - \int_{h}^{\infty}\int_{k}^{\infty} N(-\rho)dxdy \tag{2.16}$$

and

$$\int_{-h}^{\infty}\int_{-k}^{\infty} N(\rho)dxdy$$

$$= 1 - \int_{h}^{\infty} N(0,\,1)dx - \int_{k}^{\infty} N(0,\,1)dy + \int_{h}^{\infty}\int_{k}^{\infty} N(\rho)dxdy \tag{2.17}$$

where $N(\rho)$ refers to the bivariate normal distribution and $N(0, 1)$ refers to normal distribution with mean zero and variance one.

By the use of the tables and the formulas (2.11) to (2 17) we can get the probability that Z will not be less than any pre-assigned number v.

To get the probability that Z will not be greater than v we compute $\{1-F(v)\}$ by the above procedure and substract it from unity.

It may be noticed that the approach suggested in this Section, in one way, is better than the approach discussed in Sections II A and II B Geary's approach gives approximate results, though the efficiency of the approximation is guaranteed by the condition (2.2). On the other hand, Fieller's approach gives exact probabilities In case (2 2) is satisfied, the results obtained from both approaches will be approximately equal. If (2 2) is not satisfied, the procedure outlined in this section (II C) should be relied upon However, sections II A and II B have their own advantages. Section II A provides approximate distributions of x_k and y if needed, and Section II B provides quadratic equations which give probability limits directly without the help of K Pearson's tables.

III. APPLICATION TO LINEAR PROGRAMMING

A. In this section we will show an application of the above analysis. It has been said in the introductory remarks that a non-degenerate

solution of a programming problem occurs as the solution of a set of linear equations,

$$BX = Q.$$

First of all, we will show it by presenting an empirical example worked out by the author [1]. An optimum production program was computed for a model family farm in Iowa which has the inputs shown in Table 1. Five crops—corn, oats, soybeans, flax and wheat—were

TABLE 1

FIXED RESOURCES

	Land, acres		148		
	Capital, dollars		1800		
	Labor—distributed over the year as below				
	hours		*hours*		*hours*
Jan.	182	May	182	Sept.	182
Feb.	182	June	234	Oct	182
Mar.	182	July	234	Nov.	182
Apr.	182	Aug.	234	Dec.	182

considered. The relevant input-output data for the period 1928-1952 were obtained for Hancock County (Ellsworth township) in Iowa and the input coefficients were computed over the same period. Their averages were used to derive the optimum program using the techniques of linear programming [3, 5]. The matrix of the input coefficients is shown in Table 2. Land, capital and the labor in the months

TABLE 2

MATRIX OF INPUT COEFFICIENTS

	Corn	Oats	Soybeans	Flax	Wheat
Land	.02274	.02770	.05862	.09249	.09081
Capital	.31772	.27870	.70812	.96956	1.00356
Labor (May)	.05253	0	.11681	0	0
Labor (July)	.02555	.07523	.05485	.21186	.42324
Labor (August)	0	.08370	0	30910	.08650

of May, July and August were assumed to be limitational, though after deriving the program, labor requirements in other months were also checked. The objective was to maximize gross capital return and for this purpose average Iowa prices for 1952 were used. These are shown in Table 3. Thus if x_1, x_2, x_3, x_4, and x_5 represent the levels of

TABLE 3

PRICES OF CROPS USED IN THE EXAMPLE

Crop	Price Per Bushel
Corn	$1 56
Oats	0 84
Soybeans	2.79
Flax	3.81
Wheat	2.14

crops (in bushels), which may occur in the program, the problem for solution was to find the amounts x_1, x_2, x_3, x_4, and x_5 of the corn, oats, soybeans, flax, and wheat respectively which maximize the linear function:

$$(1.56)x_1 + (.84)x_2 + (2.79)x_3 + (3\ 81)x_4 + (2.14)x_5$$

subject to the inequalities

$$\text{(i)}\quad (.02274)x_1+(\ 02770)x_2+(.05862)x_3+(\ 09249)x_4+(.09081)x_5 \leqq 148$$

$$(.31772)x_1+(.27870)x_2+(.70812)x_3+(\ 96956)x_4 + (1\ 00356)x_5 \leqq 1800$$

$$(.05253)x_1+0x_2+(\ 11681)x_3+0x_4+0x_5 \leqq 182$$

$$(.02555)x_1+(\ 07523)x_2+(\ 05485)x_3+(\ 21186)x_4+(\ 42324)x_5 \leqq 234$$

$$0x_1+(\ 08370)x_2+0x_3+(.30910)x_4+(\ 08650)x_5 \leqq 234,$$

and

$$\text{(ii)}\qquad x_1,\ x_2,\ x_3,\ x_4, \text{ and } x_5 \text{ are } \geqq 0$$

Introducing the disposal variables x_6, x_7, x_8, x_9, x_{10} and using the simplex method [3, 5], the following program was derived

$$
\begin{aligned}
x_1 &= 3464.49718 \\
x_4 &= 686.73761 \\
x_6 &= 5\ 70277 \\
x_7 &= 33.44223 \\
x_{10} &= 21.73124
\end{aligned}
\tag{3.1}
$$

Thus we should plan to produce 3464 5 bushels of corn and 687 bushels of flax, which will require the inputs shown in Table 4. The labor required in this program during the rest of the months was also computed and it did not exceed the amounts available in those months.

TABLE 4

SCHEDULE OF INPUTS

Input	Unit	Corn	Flax	Disposal	Total
Land	Acres	78.78	63 52	5 70	148
Capital	Dollars	1100.73	665.83	33 44	1800
Labor					
May	Hours	182	0	0	182
July	Hours	88 51	145 49	0	234
Aug.	Hours	0	212 27	21 73	234

Now we notice that the quantities shown in (3.1) satisfy the following equations

$$
\begin{aligned}
(\ 02274)x_1 + (.09249)x_4 + (1)x_6 + (0)x_7 + (0)x_{10} &= 148 \\
(.31772)x_1 + (.96956)x_4 + (0)x_6 + (1)x_7 + (0)x_{10} &= 1800 \\
(.05253)x_1 + (0)x_4 + (0)x_6 + (0)x_7 + (0)x_{10} &= 182 \\
(\ 02555)x_1 + (\ 21186)x_4 + (0)x_6 + (0)x_7 + (0)x_{10} &= 234 \\
(0)x_1 + (.30910)x_4 + (0)x_6 + (0)x_7 + (1)x_{10} &= 234
\end{aligned}
\tag{3.2}
$$

which corresponds to the set

$$BX = Q. \tag{3.3}$$

The expected gross profit and corn flax yields are as follows.

Corn (x_1)	3464.5 bushels
Flax (x_4)	687 bushels
Gross profit (y)	\$8,021

The purpose of this section is to show a simple application of Section II. Therefore, a detailed discussion of assumptions and economic implications of the use of linear programming in farm production planing is avoided here. Interested readers may refer to [1].

B. *Statistical Analysis*. The input coefficients (the non-negative coefficients of x_1 and x_2 in equations 3 2) are the averages of the corresponding empirical values computed over the past twenty-five years The unbiased estimates of the variances computed from the same series were used to apply the analysis of Section II with the assumption of normality of the distributions of these coefficients. The fixed inputs (Q) and the prices (C) were assumed to be without random errors It can be verified that in this example, we will need only three variances to be used in formulas (1 7), (1.9), (1 10), (1.14) and (1 15) These variances are as follows:

1) Variance of the labor input coefficient for May for corn which is .000227125
2) Variance of the labor input coefficient for July for flax which is .003262968, and
3) Variance of the labor input coefficient for July for corn which is .000053646

With the help of these estimated variances and by the direct application of the quadratic equations (2 8) and (2 9), the following 95 per cent probability limits were obtained.

	Lower Limit	Upper Limit
Corn (x_1)	bushels 2,088	10,282
Flax (x_4)	bushels 46.5	1,625
Gross profit (y)	$6,298	$18,866

Comparison of the lower limit of the gross profit with its expected amount, $8021, gives an increased confidence in the program

Next the following probabilities were investigated along the lines suggested in Section II C

A. Corn yield

$$P[x_1 < 2000] = .0277$$

$$P[x_1 \geqq 10000] = .02077$$

B Flax yield

$$P[x_4 < 50] = .027$$

$$P[x_4 \geqq 1600] = 0228$$

C. Gross profit

$$P[y < 6500] = .0276$$

$$P[y \geqq 18000] = .0232$$

This means that the probability is only .0276 that the profit will be less than \$6500. The usefulness of this result is obvious.

Thus we notice that the analysis of Section II can be applied in practical problems with the advantage of easy computation. It is based on the approximate reduction of Section I, which is not objectionable if the random fluctuations in the constants involved in the model (the input coefficients in the case of linear programming) are sufficiently small.

We believe that the application of this probability approach to the problems of linear programming and inter-industry analysis gives a greater flexibility to the handling of the input coefficients.

Instead of the assumption that these input-output relations are fixed constants, not subject to any variation during the lapse of period between planning and the action, we assume that the fluctuations are possible but are random variables and normally distributed. Thus instead of estimating the fixed input coefficients, we may estimate their distributions. That is, their expected values and the variances. This may be done by using technical information or the relevant past data. The planning of the program is effected by the expected values of the input coefficients and the variances are used to form an idea of the variability of the outcome of the program

There is another situation, wherein this probability approach may be very helpful. Sometimes, we may have more than one optimum solution of a linear programming problem, which give the same value of the "objective function" The highest lower limit of the "objective function" for the same probability will then provide a criterion of choice among those alternative programs.

It is admitted that the approximation suggested in Section I is rather crude We could have retained higher order product terms of the random errors. The derivation of the distributions in that case would become highly complicated Even if the errors are normally distributed, their product is not normally distributed. For simplicity of presentation of this approach and for readily applicable results, we have confined ourselves to first order approximations only. As has been seen in the preceding example, however, and in many other practical cases, the values of the coefficients may be less than one—see eq (3.2). As a matter of fact, the unit for the variables (x_k) may be so defined that the values of the coefficients which represent inputs required per unit

of a particular output will be less than one. Therefore, the possible random errors in those coefficients will be still far smaller for practical reasons. For example, if the mean land input coefficient is .02274, the random fluctuations in it can not make it negative. Hence the random error will be far less than .02274. Neglecting their second and third order products will not introduce much serious bias in the results.

REFERENCES

[1] Babbar, M. M., *Statistical Approach in Planning Production Programs for Interdependent Activities*, Doctoral Dissertation, Ames, Iowa State College Library (1953).

[2] Box, G. E. P., and Hunter, J. S., "A confidence region for the solution of a set of simultaneous equations with an application to experimental designs," *Biometrika*, 41 (1953), 190–9.

[3] Charnes, A., Cooper, W. W., and Henderson, A., *An Introduction to Linear Programming*, New York: John Wiley and Sons, Inc. (1953).

[4] Craig, C. C., "The frequency function of y/x," *Annals of Mathematics*, 30 (1929), 471–86.

[5] Dantzig, George B., "Maximization of linear function of variables subject to linear inequalities," in *Activity Analysis of Production and Allocation*, Edited by Tjalling C. Koopmans, New York. John Wiley and Sons, Inc. (1951).

[6] Dorfman, Robert, *Application to Linear Programming to the Theory of the Firm*, Berkeley: University of California (1951).

[7] Dwyer, Paul S., *Linear Computations*, New York. John Wiley and Sons, Inc (1951).

[8] Fieller, E. C., "The biological standardization of insulin," *Journal of Royal Statistical Society*, Supplement, 7 (1940), 1–53.

[9] Fieller, E. C., "The distribution of the index in a normal bivariate population," *Biometrika*, 24 (1932), 428–40.

[10] Geary, R. C., "The frequency distribution of the quotient of two normal Variates," *Journal of the Royal Statistical Society*, 93 (1930), 442–6.

[11] Gurland, John, "Inversion formulae," *Annals of Mathematical Statistics*, 19 (1948), 228–37.

[12] Hotelling, Harold, "Some new methods in matrix calculations," *Annals of Mathematical Statistics*, 14 (1943), 1–33.

[13] Leontief, W. W., *The Structure of the American Economy*, New York: Oxford University Press (1951).

[14] Lonseth, A. T., "Systems of linear equations with coefficients subject to error," *Annals of Mathematical Statistics*, 13 (1942), 332–7.

[15] Pearson, Karl, "On the constants of index distributions, as deduced from the like constants for the components of the ratio, with special reference to the opsonic index," *Biometrika*, 7 (1910), 531–41.

[16] Pearson, Karl, *Tables for Statisticians and Biometricians, Part II*, Cambridge University Press (1924)

[17] Tintner, Gerhard, "The distribution of solutions of linear equations whose coefficients are subject to error." Unpublished, Department of Economics and Sociology, Iowa State College.

[18] Turing, A. M., "Rounding off errors in matrix processes," *Quarterly Journal of Mechanics and Applied Mathematics*, 1 (1947), 287–308.

ESTIMATION OF PARAMETERS FROM INCOMPLETE DATA

FREDERIC M. LORD
Educational Testing Service

1 INTRODUCTION AND SUMMARY

THE present note is concerned with a special case of the general problem of obtaining efficient estimators for the parameters of a normal multivariate population when the available sample data are incomplete in the sense that measures on all variables are not available for all individuals in the sample Such fragmentary data may arise because part of the data are irretrievably lost (e g , in an archaeological find), or because certain data were purposely not collected The decision not to measure all individuals in the sample on all variables may be reached because of the cost of measurement, because of limited time, because the measurement of a certain variable alters or destroys the individual measured (e.g , in mental testing, testing explosives), and so forth.

The general problem for normal bivariate populations has been treated by Wilks [3]. This treatment has been further generalized to normal multivariate populations by Matthai [2], who deals explicitly with the trivariate case. Unfortunately, the general maximum likelihood equations have proved rather intractable, and no simple formulas for the maximum likelihood estimators are available in the general case, even for a sample from a bivariate population.

In the present paper, the problem of estimating the parameters of a normal trivariate population from incomplete data is dealt with in a special case for which explicit solutions to the maximum likelihood equations are readily obtained This special case is described in the following section Formulas for the maximum likelihood estimators are given, their application is illustrated by a numerical example The sampling variances and covariances of the maximum likelihood estimators are derived An examination is made of the efficiency of the usual methods that utilize only that portion of the data that is complete.

The foregoing results are specialized to apply to a commonly encountered bivariate (rather than trivariate) situation

2. PROBLEM

We will be concerned with three variables, u, v, and w, which are assumed to have a normal trivariate distribution in the population

from which the available random sample of individuals has been drawn. In the available data, variable w is recorded for all individuals; either u or v is recorded for all individuals, but not both. The N' individuals for whom u is recorded will be denoted collectively as group a; the N'' individuals with v will be denoted as group b. The total number of individuals in the sample is $N=N'+N''$.

If u and v are correlated with w, it is obvious after a little thought that the data for group b contain some information relevant for estimating the parameters of variable u, and that the data for group a contain some information relevant for estimating the parameters of v The problem is to use the available data as efficiently as possible for estimating the parameters concerned.

3. THE MAXIMUM LIKELIHOOD ESTIMATORS

The joint distribution of u, v, and w involves nine parameters, which may be defined in various ways. A commonly used set of nine parameters is the means (μ_u, μ_v, μ_w), the variances (σ_u^2, σ_v^2, σ_w^2), and the correlations (ρ_{uw}, ρ_{vw}, ρ_{uv}). The last correlation coefficient (ρ_{uv}) will not be discussed further, since it cannot be estimated from the data at hand (Although upper and lower bounds to ρ_{uv} may be set by making use of the fact that the correlation matrix must be positive definite, these bounds are ordinarily very far apart.) For convenience, the subscript w will be dropped from the two remaining correlation coefficients.

The maximum likelihood estimators for the means are (see section 5):

$$\overset{*}{\mu}_w = \bar{w}, \tag{1}$$

$$\overset{*}{\mu}_u = \bar{u}' - \overset{*}{\beta}_u(\bar{w}' - \bar{w}), \tag{2}$$

$$\overset{*}{\mu}_v = \bar{v}'' - \overset{*}{\beta}_v(\bar{w}'' - \bar{w}). \tag{3}$$

Here the estimators are denoted by "stars" (*); $\bar{u}$, $\bar{v}$, and $\bar{w}$ are sample means, sample statistics relating to groups a and b, respectively, are distinguished from statistics relating to the combined groups by single and double primes, and $\overset{*}{\beta}_u$ and $\overset{*}{\beta}_v$ are estimators defined in equations 9 and 10 It is readily shown that $\overset{*}{\mu}_w$, $\overset{*}{\mu}_u$, and $\overset{*}{\mu}_v$ are unbiased estimators

Instead of estimating the five parameters σ_w, σ_u, σ_v, ρ_u, and ρ_v directly, it is convenient to replace these by an equivalent set consisting of σ_w, the standard errors of estimate $\sigma_{u \cdot w}$ and $\sigma_{v \cdot w}$, and the regression coefficients β_u and β_v, where

$$\left.\begin{aligned}\sigma^2_{u\cdot w} &= \sigma^2_u(1-\rho^2_u),\\ \sigma^2_{v\cdot w} &= \sigma^2_v(1-\rho^2_v),\end{aligned}\right\} \tag{4}$$

$$\left.\begin{aligned}\beta_u &= \sigma_u\rho_u/\sigma_w,\\ \beta_v &= \sigma_v\rho_v/\sigma_w.\end{aligned}\right\} \tag{5}$$

The maximum likelihood estimator for any parameter in this new set is simply the corresponding sample statistic:

$$\overset{*}{\sigma}{}^2_w = s^2_w = \frac{1}{N}\sum^N w^2 - \bar{w}^2, \tag{6}$$

$$\overset{*}{\sigma}{}^2_{u\cdot w} = s'^2_{u\cdot w} = s'^2_u(1 - r'^2_{uw}), \tag{7}$$

$$\overset{*}{\sigma}{}^2_{v\cdot w} = s''^2_{v\cdot w} = s''^2_v(1 - r''^2_{vw}), \tag{8}$$

$$\overset{*}{\beta}_u = b'_{uw} = s'_u r'_{uw}/s'_w, \tag{9}$$

$$\overset{*}{\beta}_v = b''_{vw} = s''_v r''_{vw}/s''_w, \tag{10}$$

where s, r, and b denote the usual sample standard deviation, correlation coefficient, and regression coefficient.

Maximum likelihood estimators for σ_u, σ_v, ρ_u, and ρ_v are readily obtained by solving equations 4 and 5 for these parameters and replacing all parameters by their estimators. Thus,

$$\overset{*}{\sigma}{}^2_u = \overset{*}{\sigma}{}^2_{u\cdot w} + \overset{*}{\beta}{}^2_u\overset{*}{\sigma}{}^2_w, \tag{11}$$

$$\overset{*}{\rho}{}^2_u = \frac{\overset{*}{\beta}{}^2_u\overset{*}{\sigma}{}^2_w}{\overset{*}{\sigma}{}^2_{u\cdot w} + \overset{*}{\beta}{}^2_u\overset{*}{\sigma}{}^2_w}, \tag{12}$$

and similarly for v. It should be noted that maximum likelihood estimators for σ_u, σ_v, ρ_u, and ρ_v are *not* provided by the corresponding sample statistics.

4. NUMERICAL EXAMPLE

Table 1 gives some data* used for equating two psychological tests, u and v. Tests u and w were administered to group a; tests v and w, to group b. (Tests u and v could not both be administered to the same group because of limited testing time; test w was a very short test that avoided this difficulty.) The practical problem here is to estimate

* The writer is indebted to W. H. Angoff for these data.

μ_u, μ_v, σ_u, and σ_v. (Given these estimates, it is then possible [1] to convert a score on test v to an "equivalent" score on test u, or vice versa; however, this further step does not concern us here.)

From Table 1 and equations (7–10), it is found that $\overset{*}{\beta}_u=2.309$, $\overset{*}{\beta}_v=2.060$, $\overset{*}{\sigma}{}^2_{u.w}=68.92$, $\overset{*}{\sigma}{}^2_{v.w}=58.73$. From equations (2) and (3), the required estimated means are found to be $\overset{*}{\mu}_u=37.72$ and $\overset{*}{\mu}_v=34.13$; from equation (11), it follows that the required estimated variances are $\overset{*}{\sigma}{}^2_u=162.11$ and $\overset{*}{\sigma}{}^2_v=132.91$. These results complete the solution insofar as the problem is one of estimation.

A heuristic statistical interpretation of these computations is the following. Consider the problem of estimating μ_u. It is observed from

TABLE 1

DATA FOR EQUATING TESTS u AND v

Statistic	Group a ($N'=506$)		Group b ($N''=511$)		Combined Groups
	Test u	Test w	Test v	Test w	Test w
Mean	$\bar{u}'=38.04$	$\bar{w}'=17.33$	$\bar{v}''=33.84$	$\bar{w}''=17.05$	$\bar{w}=17.19$
Variance	$s_u'^2=164.11$	$s_w'^2=17.86$	$s_v''^2=131.16$	$s_w''^2=17.06$	$s_w^2=17.48$
Correlation with w	$r'_{uw}=.7616$		$r''_{vw}=.7431$		

the data that group a has a larger sample mean in the w variable than group b, i.e., $\bar{w}'>\bar{w}''$. Since both groups are random samples from the same population, and since w is (apparently) positively correlated with u, it is plausible that the sample mean of u in group a is larger than the (unavailable) sample mean of u in group b, i.e., $\bar{u}'>\bar{u}''$, and hence $\bar{u}'>\bar{u}$, where $\bar{u}=(N'\bar{u}'+N''\bar{u}'')/N$. Since this last statistic, if it were available, would be the efficient estimate of μ_u, the estimate $\bar{u}'$ should be adjusted downward. The size of the adjustment required is given by formula (2). Similarly, for the estimation of μ_v. Analogous heuristic remarks apply to the estimation of σ_u^2 and σ_v^2. The fact that (2) and (3) provide unbiased estimators is evidence that the adjustments made in estimating μ_u and μ_v are of the correct size, even in small samples.

If the statistician did not avail himself of the relevant information in group b, he would simply use the observed mean, $\bar{u}'$, as an estimate of μ_u. The efficiency of this simple estimating procedure in the case of the present numerical example is seen from (17) to be only .71. The efficiency of $s_u'^2$ as an estimate of σ_u^2 is seen from (19) to be .83.

5. THE LIKELIHOOD FUNCTION

The detailed derivation of the formulas for the maximum likelihood estimators will be given elsewhere [1] as part of a discussion of a particular application. These formulas are derived by the usual methods from the likelihood function, which is constructed as follows.

The likelihood function of u and w for group a is clearly

$$L' = \frac{1}{(2\pi\sigma_u\sigma_w\alpha_u)^{N'}} \exp\left[-\frac{1}{2\alpha_u^2}\sum_a\left\{\frac{(u_a-\mu_u)^2}{\sigma_u^2} + \frac{(w_a-\mu_w)^2}{\sigma_w^2} - 2\rho_u\frac{(u_a-\mu_u)}{\sigma_u}\frac{(w_a-\mu_w)}{\sigma_w}\right\}\right], \qquad (13)$$

where $\alpha_u^2 = 1-\rho_u^2$ A similar expression in v and w holds for group b. The likelihood function for the entire set of data is simply the product of these two separate functions.

6. SAMPLING VARIANCES; THE EFFICIENCY OF OTHER ESTIMATORS

Table 2 gives the matrix of the large-sample variances and covariances of the maximum-likelihood estimators for the population means The estimators for β_u, β_v, $\sigma_{u\cdot w}$, $\sigma_{v\cdot w}$, and σ_w are uncorrelated in

TABLE 2

MATRIX OF THE SAMPLING VARIANCES AND COVARIANCES OF THE MAXIMUM LIKELIHOOD ESTIMATORS OF THE POPULATION MEANS

	$\overset{*}{\mu}_u$	$\overset{*}{\mu}_w$	$\overset{*}{\mu}_v$
$\overset{*}{\mu}_u$	$\frac{\sigma_u^2}{N}\left(1+\frac{N''}{N'}\alpha_u^2\right)$	$\frac{1}{N}\sigma_u\sigma_w\rho_u$	$\frac{1}{N}\sigma_u\sigma_v\rho_u\rho_v$
$\overset{*}{\mu}_w$	$\frac{1}{N}\sigma_u\sigma_w\rho_u$	$\frac{\sigma_w^2}{N}$	$\frac{1}{N}\sigma_v\sigma_w\rho_v$
$\overset{*}{\mu}_v$	$\frac{1}{N}\sigma_u\sigma_v\rho_u\rho_v$	$\frac{1}{N}\sigma_v\sigma_w\rho_v$	$\frac{\sigma_v^2}{N}\left(1+\frac{N'}{N''}\alpha_v^2\right)$

large samples; the large-sample variances of these estimators are as follows:

$$\text{Var } \overset{*}{\sigma}{}_w^2 = \frac{2\sigma_w^4}{N}, \tag{14}$$

$$\text{Var } \overset{*}{\sigma}{}_{u\cdot w}^2 = \frac{2}{N'}\,\sigma_u^4\alpha_u^4, \tag{15}$$

$$\text{Var } \overset{*}{\beta}_u = \frac{\sigma_u^2\alpha_u^2}{N'\sigma_w^2}, \tag{16}$$

and so forth

Various investigations can be made regarding the amount of additional information obtained by making full use of the fragmentary data available. For example, if μ_u is estimated by $\bar{u}'$, the sampling variance of the estimate is, of course, σ_u^2/N'. The efficiency of this estimate in the present situation is, of course, the reciprocal of the ratio of its sampling variance to the sampling variance in the upper left cell of Table 2, viz.,

$$1 - \frac{N''}{N}\,\rho_u^2. \tag{17}$$

It is apparent that if ρ_u is large and the proportion of cases in the b-group is large, then utilization of the data on w in the b-group will improve the estimation of μ_u very considerably

Equations (7) and (9) show that estimates of β_u and $\sigma_{u\cdot w}^2$ are not improved by using information from the b-group data. Estimates of σ_u and of ρ_u will be improved, however. For example, the sampling variance of $\overset{*}{\sigma}{}_u^2$ in large samples is found from (11), (12), (14), (15), (16) to be

$$\text{Var } \overset{*}{\sigma}{}_u^2 = \frac{2\sigma_u^4}{N'}\left(1 - \frac{N''}{N}\,\rho_u^4\right). \tag{18}$$

The efficiency of $s_u'^2$ as an estimate of σ_u^2 is seen to be

$$1 - \frac{N''}{N}\,\rho_u^4. \tag{19}$$

Other formulas may be similarly obtained for other parameters.

7. THE BIVARIATE CASE

It is seen that the observed values of variable v do not enter into (1), (2), (6), (7), (9) Consequently, the formulas derived here for u and w apply without change to the bivariate case of incomplete data where one variable is observed for all individuals in the sample but the second variable is not. In other words, if complete data on w are available for $N'+N''$ cases whereas data on u are available on only N' of these cases, then the maximum-likelihood estimates of μ_u, $\sigma_{u \cdot w}$, and β_u are given by equations (2), (7) ,and (9), as before.

REFERENCES

[1] Lord, F., "Equating test scores—a maximum likelihood solution," *Psychometrika*, in press

[2] Matthai, A., "Estimation of parameters from incomplete data with application to design of sample surveys," *Sankhyā*, 11 (1951), 145–52.

[3] Wilks, S., "Moments and distributions of estimates of population parameters from fragmentary samples," *Annals of Mathematical Statistics*, 3 (1932), 163–203.

TRUNCATED BINOMIAL AND NEGATIVE BINOMIAL DISTRIBUTIONS

PAUL R. RIDER
Wright-Patterson Air Force Base

INTRODUCTION

CONSIDERABLE interest has recently been manifested in truncated distributions. (See references at end of paper.) Finney [10] has treated the truncated binomial distribution and has shown how to estimate the parameter of the distribution by an iterative process which requires special tables. He mentions several practical problems in which a truncated binomial distribution might be met. The first part of the present paper shows how to estimate the parameter by a simple method analogous to that previously used [16] in estimating the parameter of a truncated Poisson distribution. The second part of the paper uses the same method to develop formulas for estimating the parameters of a truncated negative binomial distribution.

THE POSITIVE BINOMIAL DISTRIBUTION

If the probability of happening of an event in a single trial is p, the probability that it will happen exactly x times in n trials is

$$p_x = \frac{n!}{x!(n-x)!} p^x(1-p)^{n-x}; \quad x = 0, 1, \cdots, n. \tag{1}$$

The expression (1) is called the binomial distribution function, since it is the general term in the binomial expansion of $(q+p)^n$, in which $q=1-p$.

A common problem is that of estimating the probability p from a set of data obtained by observing the results of an experiment. The problem with which we are concerned here is that of estimating p from a sample taken from a truncated distribution.

Let us designate by f_x the frequency with which the value x occurs in the sample. Then the expected value of f_x is Np_x, N being the unknown total of the untruncated sample. Suppose that k classes are truncated from the lower end of the distribution.[1] We shall use the following notation, in which x is the index of summation:

[1] If the truncation occurs at the upper end of the distribution the roles of p and q are simply interchanged.

$$T_0 = \sum_k^n f_x, \qquad T_1 = \sum_k^n x f_x, \qquad T_2 = \sum_k^n x^2 f_x, \tag{2}$$

$$T_0' = N \sum_0^{k-1} p_x + T_0, \tag{3}$$

$$T_1' = N \sum_0^{k-1} x p_x + T_1, \tag{4}$$

$$T_2' = N \sum_0^{k-1} x^2 p_x + T_2. \tag{5}$$

Then T_1'/T_0' is an estimate of the first moment of the distribution, namely np, and T_2'/T_1' is an estimate of the second moment about the origin, $np(q+np)$, divided by the first moment, this quotient being $q+np$. We therefore set $T_1' = npT_0'$, $T_2' = (q+np)T_1'$, that is,

$$N \sum_0^{k-1} \frac{n!}{x!(n-x)!} x p^x q^{n-x} + T_1 = Nnp \sum_0^{k-1} \frac{n!}{x!(n-x)!} p^x q^{n-x} + npT_0, \tag{6}$$

$$N \sum_0^{k-1} \frac{n!}{x!(n-x)!} x^2 p^x q^{n-x} + T_2 = N(q+np) \sum_0^{k-1} \frac{n!}{x!(n-x)!} x p^x q^{n-x} + (q+np)T_1. \tag{7}$$

Then we solve (6) for q^{n-k+1}, obtaining

$$q^{n-k+1} = \frac{(k-1)!(T_1 - npT_0)}{Nn(n-1)\cdots(n-k+1)p^k} \tag{8}$$

Next we substitute this value in (7) and solve for p The resulting estimate is

$$p = \frac{T_2 - kT_1}{(n-1)T_1 - (k-1)nT_0}. \tag{9}$$

The estimate of N can be obtained from the equation

$$N - T_0 = N \sum_0^{k-1} \frac{n!}{x!(n-x)!} p^x q^{n-x}. \tag{10}$$

THE NEGATIVE BINOMIAL DISTRIBUTION

The negative binomial distribution function is [12]

$$p_x = \frac{(m+x-1)!}{x!(m-1)!} \frac{p^x}{(1+p)^{m+x}} \tag{11}$$

This is the general term in the binomial expansion of $(r-p)^{-k}$, in which $r=1+p$. As we wish to estimate p and m, also the number N in a sample before truncation, we need to use three moments. The first three moments about the origin are

$$\mu_1' = mp, \qquad \mu_2' = mp(1 + p + mp),$$
$$\mu_3' = mp(1 + 3p + 2p^2 + 3mp + 3mp^2 + m^2p^2). \tag{12}$$

We shall consider only the case in which the class corresponding to $x=0$ has been truncated, formulas for the general case are too complicated to be of interest. In the notation of formulas (2)–(5) with $k=1$, we have, since $p_0=(1+p)^{-m}$,

$$T_0' = N(1 + p)^{-m} + T_0; \quad T_i' = T_i, \ i > 0. \tag{13}$$

Now T_2'/T_1' is an estimate of μ_2'/μ_1', and consequently we set

$$T_2 = T_1(1 + p + mp), \tag{14}$$

$$T_3 = T_1(1 + 3p + 2p^2 + 3mp + 3mp^2 + m^2p^2). \tag{15}$$

Next we solve (13) for mp and substitute in (14)
Solving the resulting equation, we find the estimates

$$p = \frac{T_3T_1 - T_2T_1 + T_1^2 - T_2^2}{T_1(T_2 - T_1)}, \tag{16}$$

$$r = 1 + p = \frac{T_3T_1 - T_2^2}{T_1(T_2 - T_1)} \cdot \tag{17}$$

From (14) and (17) we obtain the estimate

$$m = \frac{2T_2^2 - T_2T_1 - T_3T_1}{T_3T_1 - T_2T_1 + T_1^2 - T_2^2} \tag{18}$$

An estimate of N can now be obtained from the equation

$$N - T_0 = Nr^{-m}. \tag{19}$$

COMPARISON WITH MAXIMUM LIKELIHOOD ESTIMATES

In this section the proposed estimators aré compared with maximum likelihood estimators

For the truncated binomial distribution the maximum likelihood estimate of p is given by the solution of the equation

$$\nu_1^{(k)} = \frac{np - \sum_0^{k-1} \frac{n!x}{x!(n-x)!} p^x(1-p)^{n-x}}{1 - \sum_0^{k-1} \frac{n!}{x!(n-x)!} p^x(1-p)^{n-x}}, \tag{20}$$

in which the left side is the first moment of the sample from the truncated distribution, k being the number of classes truncated. (Cf [16].) Special cases of (20) are

$$\nu_1' = \frac{np}{1-(1-p)^n}, \tag{21}$$

$$\nu_1'' = \frac{np - np(1-p)^{n-1}}{1-(1-p)^n - np(1-p)^{n-1}}. \tag{22}$$

For the purpose of comparison we use Table 1, taken from [11, sec 18]. This table exhibits data concerning 53,680 families having eight children each The number of boys in each such family is denoted by x, the corresponding frequency by f. The mean number of boys is 4 117418; the value of p is one eighth of this, or 0 514677.

Using the estimator proposed (as if the frequencies of certain classes were unknown), we find:

For $k=1$; $T_0=53680-215=53465$, $T_1=221023$, $T_2=1021023$, and, from (9), $p=0\ 517$. The estimate of N, obtained by using (10) is 53624.

TABLE 1

NUMBERS OF BOYS IN FAMILIES HAVING EIGHT CHILDREN

Number of Boys x	Number of Families f	xf	x^2f
0	215	0	0
1	1485	1485	1485
2	5331	10662	21324
3	10649	31947	95841
4	14959	59836	239344
5	11929	59645	298225
6	6678	40068	240408
7	2092	14644	102508
8	342	2736	21888
Total	53680	221023	1021023

TABLE 2

VALUES OF ν_1'

p \ n	2	3	4	5	6	7	8	9	10
0.1	1 053	1.107	1.163	1.221	1.281	1 342	1 405	1.469	1.535
0.2	1 111	1 230	1 355	1 487	1 626	1.772	1 923	2 079	2.241
0.3	1 176	1 370	1 579	1 803	2 040	2.288	2.547	2 814	3.087
0 4	1.250	1 531	1.838	2.169	2 517	2 881	3.255	3.637	4.024
0.5	1.333	1 714	2 133	2 581	3.058	3.528	4 016	4.509	5 005
0 6	1 429	1 923	2 463	3 031	3 615	4 207	4.803	5.401	6 000
0.7	1 538	2 158	2.823	3.509	4.206	4 901	5.600	6.300	7.000
0 8	1 667	2 419	3 205	4 001	4.800	5.600	6 400	7.200	8.000
0.9	1.818	2 703	3 600	4 500	5 400	6 300	7 200	8 100	9.000

For $k=2$; $T_0=53680-215-1485=51980$, $T_1=221023-1485=219538$, $T_2=1021023-1485=1019538$ and, from (9), $p=0\ 518$. From (10), our estimate of N is 53475

In obtaining a maximum likelihood estimate of p from a binomial distribution from which one class has been truncated ($k=1$), a table such as Table 2 is convenient. In the present example $\nu_1'=221023/53465=4.1340$ In the column for $n=8$ we interpolate between 4 016 and 4 803 and find $p=0.515$, a correct estimate

When $k=2$ the situation is much more complicated. Equation (22) must be solved In the illustrative example, $\nu_1''=219538/51980=4\ 223509$, and (22) becomes

$$4\ 223509 = \frac{8p - 8p(1-p)^7}{1-(1-p)^8 - 8p(1-p)^7}.$$

It is convenient to replace p by $1-q$ Making this substitution and reducing, we are led to the equation

$$q^8 - 1.195609q^7 + 0.370903q - 0\ 175089 = 0.$$

Now this equation, although of eighth degree, is not too difficult to solve, since so many powers of the unknown are missing. It has the root $q=0\ 485$, giving $p=0.515$, again a correct estimate. However, the work entailed in obtaining it, as compared with that required to obtain the estimate by the simple method suggested in this paper, seems hardly to justify the slight gain in accuracy. (It is recalled that the estimate referred to was 0 518)

For illustrating the estimators of the parameters of a truncated negative binomial distribution we use Table 3, taken from [2, p. 186]

TABLE 3

NUMBERS (x) OF YEAST CELLS PER SQUARE IN A HEMOCYTOMETER

x	f	xf	2_f	3_f
0	213	0	0	0
1	128	128	128	128
2	37	74	148	296
3	18	54	162	486
4	3	12	48	192
5	1	5	25	125
Total	400	273	511	1227

In this table, x is the number of yeast cells per square in a hemocytometer, f is the corresponding observed frequency in a certain experiment. As well as can be ascertained, the estimates used by the authors are $p=0\ 192276$, $m=3\ 549593$; they purport to have been determined by the method of maximum likelihood

If the frequency corresponding to $x=0$ were unknown, then in the notation given earlier we should have $T_0=400-213=187$, $T_1=273$, $T_2=511$, $T_3=1227$ The estimate of r, as given by (17), is

$$r = \frac{(1227)(273) - (511)^2}{273(511 - 273)} = 1\ 136608,$$

from which we get $p=0\ 136608$. The estimate of m, as given by (18), is

$$m = \frac{2(511)^2 - (511)(273) - (1227)(273)}{(1227)(273) - (511)(273) + (273)^2 - (511)^2} = 5\ 381703.$$

From (19) we obtain the estimate $N=376$ These estimates are reasonably good, especially when it is taken into consideration that over half of the observations are in the zero class.

Maximum likelihood equations for estimating the parameters of an untruncated negative binomial distribution are given in [2] Since they are extremely difficult to solve and since the corresponding equations for a truncated distribution would be worse, no investigation of such equations has been made in this study.

REFERENCES

[1] Bliss, C. I., "Estimation of the mean and its error from incomplete Poisson distributions," *Connecticut Agricultural Experiment Station Bulletin*, 513 (1948), 12 pp

[2] Bliss, C. I, and Fisher, R A., "Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial," *Biometrics*, 9 (1953), 176–200

[3] Cochran, W. G., "Use of IBM equipment in an investigation of the truncated normal problem," *Proceedings of Research Forum*, International Business Machines Corp , 1946, 40–3.

[4] Cohen, A C , Jr , "On estimating the mean and standard deviation of truncated normal distributions," *Journal of the American Statistical Association*, 44 (1949), 518–25.

[5] Cohen, A C., Jr , "Estimating parameters of Pearson type III populations from truncated samples," *Journal of the American Statistical Association*, 45 (1950), 411–23.

[6] Cohen, A. C., Jr., "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples," *Annals of Mathematical Statistics*, 21 (1950), 557–69

[7] Cohen, A. C , Jr , "Estimation of parameters in truncated Pearson frequency distributions," *Annals of Mathematical Statistics*, 22 (1951), 255–65

[8] Cohen, A. C , Jr , "Estimation of the Poisson parameter from truncated samples and from censored samples," *Journal of the American Statistical Association*, 49 (1954), 158–68.

[9] David, F. N., and Johnson, N L , "The truncated Poisson," *Biometrics*, 8 (1952), 275–85

[10] Finney, D J., "The truncated binomial distribution," *Annals of Eugenics*, 14 (1949), 319–28

[11] Fisher, R. A., *Statistical Methods for Research Workers*, London. Oliver and Boyd, 11th Edition, 1940

[12] Fisher, R A , "The negative binomial distribution," *Annals of Eugenics*, 11, part 2 (1941), 182–7 This paper is reprinted as paper 38 in Fisher, R. A , *Contributions to Mathematical Statistics*, New York John Wiley and Sons, Inc , 1950

[13] Hald, A , "Maximum likelihood estimation of the parameters of a normal distribution which is truncated at a known point," *Skandinavisk Aktuarietidskrift*, Haft 3–4 (1949), 119–34

[14] Ipsen, Johannes, Jr , "A practical method of estimating the mean and standard deviation of truncated normal distributions," *Human Biology*, 21 (1949), 1–16.

[15] Moore, P G., "The estimation of the Poisson parameter from a truncated distribution," *Biometrika*, 39 (1952), 247–51

[16] Rider, Paul R , "Truncated Poisson distributions," *Journal of the American Statistical Association*, 48 (1953), 826–30

[17] Stevens, W L , "The truncated normal distribution," Appendix to "The calculation of the time-mortality curve" by C I Bliss, *Annals of Applied Biology*, 24 (1937), 815–52

RESTRICTION AND SELECTION IN SAMPLES FROM BIVARIATE NORMAL DISTRIBUTIONS*

A. CLIFFORD COHEN, JR.
The University of Georgia

1. INTRODUCTION AND SUMMARY

THE problem considered here is that of estimating parameters of a bivariate normal population with probability density (frequency) function

$$f(x, y) = \tag{1} \frac{\exp \dfrac{-1}{2(1-\rho^2)}\left[\left(\dfrac{x-m_x}{\sigma_x}\right)^2 - 2\rho\left(\dfrac{x-m_x}{\sigma_x}\right)\left(\dfrac{y-m_y}{\sigma_y}\right) + \left(\dfrac{y-m_y}{\sigma_y}\right)^2\right]}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}},$$

from restricted samples of types which arise when acceptance or screening procedures based on one variate, eliminate certain sample specimens from further observation with respect to the other variate. This problem is important in connection with correlation studies which relate entrance examination scores to subsequent achievement scores and in numerous similar situations. In analyses of acceptance inspection data, it becomes important when a physical characteristic such as weight, density, size, hardness, etc., must be correlated with a performance characteristic such as life span, operating cost, sales volume or other characteristic for which observations are available only on accepted items.

Without loss of generality, we let x be the restricted variate, and in order to emphasize the dependence of y on x, we write $f(x, y)$ as the product of the marginal frequency function of x and the conditional (array) frequency function of y. Thus (1) becomes

$$\tag{2} \begin{aligned} f(x, y) = &\left\{ \frac{\exp -\dfrac{1}{2}\left[\dfrac{x - m_x}{\sigma_x}\right]^2}{\sigma_x\sqrt{2\pi}} \right\} \\ &\times \left\{ \frac{\exp -\dfrac{1}{2}\left[\dfrac{y - \alpha - \beta(x - \bar{x})}{\sigma}\right]^2}{\sigma\sqrt{2\pi}} \right\}, \end{aligned}$$

where

$$\tag{3} \beta = \rho\sigma_y/\sigma_x, \quad \alpha = m_y - \beta(m_x - \bar{x}), \quad \sigma^2 = \sigma_y^2(1 - \rho^2),$$

* Sponsored by the Office of Ordnance Research, U. S. Army, under contract DA-01-009-ORD-288.

and for a given sample with n accepted specimens, $\bar{x}$ is the x-mean of the accepted specimens only ($\bar{x}=\Sigma_1^n x_i/n$). The introduction of $\bar{x}$ into (2) is for the purpose of insuring the mutual independence of estimates of α, β, and σ as is noted in Section 3. Regardless of the value of x, the corresponding y-array distribution is normal with standard deviation, σ, and with mean, $\alpha+\beta(x-\bar{x})$.

The method of maximum likelihood is employed to estimate parameters of (2), and asymptotic variances and covariances are obtained. Estimators (estimates) of m_y, σ_y, and ρ are then obtained from (3). To demonstrate the practical application of results obtained here, an illustrative example is included

Results somewhat closely related to the problem considered here have been previously published by various authors including Pearson [9], Aitken [1], Wilks [11], Lawly [8], Birnbaum [2], Birnbaum, Paulsen and Andrews [3], Campbell [4], Des Raj [10], and the writer [5, 6].

2. MAXIMUM LIKELIHOOD ESTIMATION

Consider the random selection of N sample specimens from a population distributed according to (1). Each of these specimens is either measured or censored with respect to x, or is entirely eliminated from observation. Of the specimens measured with respect to x, those which meet acceptance criteria are also measured with respect to y. Acceptance criteria may or may not be the same as those which govern observation with respect to x Thus, each specimen measured with respect to y is also measured with respect to x, but the converse of this statement is not necessarily true. We let $n(\leqq N)$ designate the number of "accepted" sample specimens and therefore the number of paired observations (x, y). Using (2), the likelihood function for a sample of this type may be put into the form

$$
\begin{aligned}
P = kG(m_x, \sigma_x)&\left\{\frac{\exp -\dfrac{1}{2}\displaystyle\sum_1^n\left(\frac{x_i - m_x}{\sigma_x}\right)^2}{(\sigma_x\sqrt{2\pi})^n}\right\} \\
&\times\left\{\frac{\exp -\dfrac{1}{2}\displaystyle\sum_1^n\left[\frac{y_i - \alpha - \beta(x_i - \bar{x})}{\sigma}\right]^2}{(\sigma\sqrt{2\pi})^n}\right\},
\end{aligned}
\tag{4}
$$

where k is a constant and $G(m_x, \sigma_x)$ is a restriction function which depends on the restrictions imposed with respect to observation of x,

and on the acceptance criteria. The function $G(m_x, \sigma_x)$ also may depend on the restricted values of x. For an unrestricted (complete) sample,

$$G(m_x, \sigma_x) \equiv 1, \qquad \text{with } n = N. \tag{5}$$

To estimate m_x, σ_x, α, β, and σ, we take logarithms of (4), differentiate and equate to zero With L written for log P, we thereby obtain

$$
\begin{aligned}
\frac{\partial L}{\partial m_x} &= \frac{1}{\sigma_x}\sum_1^n\left(\frac{x_i - m_x}{\sigma_x}\right) + \frac{1}{G}\,\frac{\partial G}{\partial m_x} = 0,\\
\frac{\partial L}{\partial \sigma_x} &= \frac{1}{\sigma_x}\sum_1^n\left(\frac{x_i - m_x}{\sigma_x}\right)^2 - \frac{n}{\sigma_x} + \frac{1}{G}\,\frac{\partial G}{\partial \sigma_x} = 0,\\
\frac{\partial L}{\partial \alpha} &= \sum_1^n\,[y_i - \alpha - \beta(x_i - \bar{x})]/\sigma^2 = 0,\\
\frac{\partial L}{\partial \beta} &= \sum_1^n\,[y_i - \alpha - \beta(x_i - \bar{x}][x_i - \bar{x}]/\sigma^2 = 0,\\
\frac{\partial L}{\partial \sigma} &= \sum_1^n\,[y_i - \alpha - \beta(x_i - \bar{x}]^2/\sigma^3 - n/\sigma = 0.
\end{aligned} \tag{6}
$$

The last three equations of (6) are independent of m_x and σ_x, and are therefore unaffected by $G(m_x, \sigma_x)$. Consequently, regardless of the acceptance criteria or of the type of restriction imposed on observation of x, these are the familiar "normal" equations, and have the well-known solutions

$$\hat{\beta} = \bar{r}\bar{s}_y/\bar{s}_x, \qquad \hat{\alpha} = \bar{y}, \quad \text{and} \quad \hat{\sigma} = \bar{s}_y\sqrt{1 - \bar{r}^2}, \tag{7}$$

where ( ) serves to distinguish maximum likelihood estimates from parameters estimated and the bars ($^{-}$) indicate that the statistics thus designated are computed solely from observations made on the n accepted sample specimens. Accordingly, $\bar{x} = \sum_1^n x_i/n$, $\bar{y} = \sum_1^n y_i/n$, $\bar{s}_x^2 = \sum_1^n (x_i - \bar{x})^2/n$, $\bar{s}_y^2 = \sum_1^n (y_i - \bar{y})^2/n$, and $\bar{r} = \sum_1^n (x_i - \bar{x})(y_i - \bar{y})/n\bar{s}_x\bar{s}_y$. Estimators (estimates) $\hat{m}_y$, $\hat{\sigma}_y$, and $\hat{\rho}$ follow from (3) and (7) as

$$
\begin{aligned}
\hat{m}_y &= \hat{\alpha} + \hat{\beta}(\hat{m}_x - \bar{x}),\\
\hat{\sigma}_y &= \sqrt{\hat{\sigma}^2 + \hat{\sigma}_x^2\hat{\beta}^2},\\
\hat{\rho} &= \hat{\sigma}_x\hat{\beta}/\sqrt{\hat{\sigma}^2 + \hat{\sigma}_x^2\hat{\beta}^2},
\end{aligned} \tag{8}
$$

and in the equivalent forms

$$
\begin{aligned}
\hat{m}_y &= \bar{y} - \bar{r}(\bar{s}_y/\bar{s}_x)(\bar{x} - \hat{m}_x), \\
\hat{\sigma}_y &= \bar{s}_y\sqrt{[1 - \hat{\lambda}(1 - \bar{r}^2)]/(1 - \hat{\lambda})}, \qquad (9)\\
\hat{\rho} &= \bar{r}/\sqrt{1 - \hat{\lambda}(1 - \bar{r}^2)},
\end{aligned}
$$

where $\hat{\lambda} = 1 - \bar{s}_x^2/\hat{\sigma}_x^2$

Estimators given in (9) above, were obtained earlier in reference [6] by setting up the likelihood function directly from (1) rather than from (2) as has been done here It was then maximized with respect to m_x, σ_x, m_y, σ_y, and ρ without the introduction of α, β, and σ. Similar results were obtained independently at about the same time by Des Raj [10]

The first two equations of (6) are simply the estimating equations for restricted samples from univariate normal populations in a form that differs only slightly from corresponding equations given in reference [5] In practice, therefore, it is merely necessary that parameters of the x-marginal distribution be estimated as in the univariate cases. With $\hat{m}_x$ and $\hat{\sigma}_x$ thus determined, $\hat{m}_y$, $\hat{\sigma}_y$, and $\hat{\rho}$ follow from (7) and (8) or from (9).

Singly Truncated Samples. When the sampling procedure is such that selection is continued until n specimens for which $x \geqq x_0$ have been measured with respect to x, accepted, and subsequently measured with respect to y, and when moreover, it is not possible to observe specimens for which $x < x_0$, so that the number of eliminated observations is unknown, the sample is said to be singly truncated on the left at terminal x_0, with respect to the x-marginal distribution. The restriction function for a sample of this type is

$$G(m_x, \sigma_x) = [I_0(\xi)]^{-n}, \tag{10}$$

where

$$I_0(\xi) = \int_\xi^\infty \phi(t)dt, \quad \xi = (x_0 - m_x)/\sigma_x, \quad \phi(t) = (\sqrt{2\pi})^{-1} \exp -t^2/2. \tag{11}$$

As is well known (c.f. [5]), estimating equations for $\hat{m}_x$ and $\hat{\sigma}_x$, in this case, become

$$
\begin{aligned}
&\text{a. } [1 - \xi(Z - \xi)]/(Z - \xi)^2 - \nu_2/\nu_1^2 = 0, \\
&\text{b. } \hat{\sigma}_x = \nu_1/(\hat{Z} - \hat{\xi}), \qquad (12)\\
&\text{c. } \hat{m}_x = x_0 - \hat{\sigma}_x\hat{\xi},
\end{aligned}
$$

where $Z(\xi)=\phi(\xi)/I_0(\xi)$ and $\nu_k=\sum_1^n(x_i-x_0)^k/n$. With ν_1 and ν_2 calculated from sample data, $\hat{\xi}$ is obtained as the solution of (12a) and $\hat{\sigma}_x$ and $\hat{m}_x$ follow from (12b) and (12c). Tables of the first member of (12a) (multiplied by $\frac{1}{2}$) and of $1/(Z-\xi)$, which are available in reference [7], greatly facilitate computation of these estimates.

Singly Censored Samples. When the sampling procedure is such that N sample specimens are selected from a population distributed according to (1), and measurements of both x and y are recorded if $x\geqq x_0$, while a count is kept of censored specimens for which $x<x_0$, although no further measurements are made, the resulting sample is of the singly censored type. The x-marginal sample data thus consist of n measured observations for which $x\geqq x_0$, and the information that $n_1=(N-n)$ sample specimens were selected for which $x<x_0$. The estimating equations in this case differ from (12) for the truncated case only in that $Z(\xi)$ in (12) is replaced with $Y(\xi)$ where

$$Y(\xi) = (n_1/n)\{\phi(\xi)/[1 - I_0(\xi)]\} = (n_1/n)Z(-\xi). \tag{13}$$

Selected Samples When the sampling procedure is such that full measurement is made and recorded with respect to x for each of the N sample specimens although corresponding y values are determined only for the n accepted specimens, the restriction function can be expressed as

$$G(m_x, \sigma_x) = (\sigma_x\sqrt{2\pi})^{n-N} \exp -\frac{1}{2}\sum_1^{N-n}[(x_i - m_x)/\sigma_x]^2. \tag{14}$$

In this case, the first two equations of (6) become

$$\begin{aligned} &\frac{1}{\sigma_x}\sum_1^N\left(\frac{x_i - m_x}{\sigma_x}\right) = 0, \\ &\frac{1}{\sigma_x}\sum_1^N\left(\frac{x_i - m_x}{\sigma_x}\right)^2 - \frac{N}{\sigma_x} = 0. \end{aligned} \tag{15}$$

From these, we obtain the usual estimates

$$\hat{m}_x = \bar{x}_N = \sum_1^N x_i/N, \qquad \hat{\sigma}_x = s_x = \sqrt{\sum_1^N (x_i - \bar{x}_N)^2/N}, \tag{16}$$

where $\bar{x}_N$ is the x-mean of the entire N sample observations and is to be distinguished from $\bar{x}$ which is used elsewhere in this paper to designate the mean of the n accepted sample specimens.

Doubly Restricted Samples. When a double restriction is imposed on observation of x at terminals x_0 and x_0+w, such that measurements are limited to specimens for which $x_0 \leq x \leq x_0+w$, the x-marginal samples are doubly truncated or doubly censored according to whether the number of unmeasured specimens is unknown or known. With samples of these types, the univariate estimating equations given in reference [5] are applicable for determining $\hat{m}_x$ and $\hat{\sigma}_x$

3. RELIABILITY OF ESTIMATES

The variance-covariance matrix of $(\hat{m}_x, \hat{\sigma}_x, \hat{\alpha}, \hat{\beta}, \hat{\sigma})$ can be derived from second order derivatives of L. Differentiating (6), we obtain

$$\frac{\partial^2 L}{\partial m_x^2} = -\frac{n}{\sigma_x^2} + \frac{1}{G}\frac{\partial^2 G}{\partial m_x^2} - \left(\frac{1}{G}\frac{\partial G}{\partial m_x}\right)^2,$$

$$\frac{\partial^2 L}{\partial m_x \partial \sigma_x} = -\frac{2}{\sigma_x^2}\sum_1^n\left(\frac{x_i - m_x}{\sigma_x}\right) + \frac{1}{G}\frac{\partial^2 G}{\partial m_x \partial \sigma_x} - \left(\frac{1}{G}\frac{\partial G}{\partial m_x}\right)\left(\frac{1}{G}\frac{\partial G}{\partial \sigma_x}\right),$$

$$\frac{\partial^2 L}{\partial \sigma_x^2} = -\frac{3}{\sigma_x^2}\sum_1^n\left(\frac{x_i - m_x}{\sigma_x}\right)^2 + \frac{n}{\sigma_x^2} + \frac{1}{G}\frac{\partial^2 G}{\partial \sigma_x^2} - \left(\frac{1}{G}\frac{\partial G}{\partial \sigma_x}\right)^2,$$

$$\text{(17)} \qquad \frac{\partial^2 L}{\partial \alpha^2} = -\frac{n}{\sigma^2}, \qquad \frac{\partial^2 L}{\partial \beta^2} = -\sum_1^n\left(\frac{x_i - \bar{x}}{\sigma}\right)^2,$$

$$\frac{\partial^2 L}{\partial \sigma^2} = -\frac{3}{\sigma^2}\sum_1^n\left[\frac{y_i - \alpha - \beta(x_i - \bar{x})}{\sigma}\right]^2 + \frac{n}{\sigma^2},$$

$$\frac{\partial^2 L}{\partial m_x \partial \alpha} = 0, \quad \frac{\partial^2 L}{\partial m_x \partial \beta} = 0, \quad \frac{\partial^2 L}{\partial m_x \partial \sigma} = 0, \quad \frac{\partial^2 L}{\partial \sigma_x \partial \alpha} = 0, \quad \frac{\partial^2 L}{\partial \sigma_x \partial \beta} = 0,$$

$$\frac{\partial^2 L}{\partial \sigma_x \partial \sigma} = 0, \quad \frac{\partial^2 L}{\partial \alpha \partial \beta} = -\frac{1}{\sigma^2}\sum_1^n (x_i - \bar{x}),$$

$$\frac{\partial^2 L}{\partial \alpha \partial \sigma} = -\frac{2}{\sigma^2}\sum_1^n [y_i - \alpha - \beta(x_i - \bar{x})]/\sigma,$$

$$\frac{\partial^2 L}{\partial \beta \partial \sigma} = -\frac{2}{\sigma^2}\sum_1^n\left[\frac{y_i - \alpha - \beta(x_i - \bar{x})}{\sigma}\right][x_i - \bar{x}].$$

To determine expected values of the terms of (17), we require the following results:

$$
(18)\quad
\begin{aligned}
&E\left[\frac{y-\alpha-\beta(x-\bar{x})}{\sigma}\right]\\
&\qquad = \frac{1}{E(n/N)}\int_R \phi(\mu)d\mu \int_{-\infty}^{\infty} v\phi(v)dv = 0,\\
&E\left[\frac{y-\alpha-\beta(x-\bar{x})}{\sigma}\right]^2\\
&\qquad = \frac{1}{E(n/N)}\int_R \phi(\mu)d\mu \int_{-\infty}^{\infty} v^2\phi(v)dv = 1,\\
&E\left[\left(\frac{y-\alpha-\beta(x-\bar{x})}{\sigma}\right)(x-\bar{x})\right]\\
&\qquad = \frac{1}{E(n/N)}\int_R (x-\bar{x})\phi(\mu)d\mu \int_{-\infty}^{\infty} v\phi(v)dv = 0,
\end{aligned}
$$

where $u=(x-m_x)/\sigma_x$, $v=[y-\alpha-\beta(x-\bar{x})]/\sigma$, and R designates the acceptance interval with respect to x. Using these results and taking expected values of the derivatives of (17), the non-vanishing terms become

$$
(19)\quad
\begin{aligned}
&-E\left[\frac{\partial^2 L}{\partial m_x^2}\right] = \frac{\phi_{11}}{\sigma_x^2}E(n^*), &\quad& -E\left[\frac{\partial^2 L}{\partial m_x \partial \sigma_x}\right] = \frac{\phi_{12}}{\sigma_x^2}E(n^*),\\
&-E\left[\frac{\partial^2 L}{\partial \sigma_x^2}\right] = \frac{\phi_{22}}{\sigma_x^2}E(n^*), && -E\left[\frac{\partial^2 L}{\partial \alpha^2}\right] = \frac{1}{\sigma^2}E(n),\\
&-E\left[\frac{\partial^2 L}{\partial \beta^2}\right] = \frac{\bar{\sigma}_x^2}{\sigma^2}E(n), && -E\left[\frac{\partial^2 L}{\partial \sigma^2}\right] = \frac{2}{\sigma^2}E(n).
\end{aligned}
$$

Expected values of all derivatives of (17) not shown in (19) are zeros We take note of the fact that $\partial^2 L/\partial\alpha\partial\beta=0$, $\partial^2 L/\partial\alpha\partial\sigma=0$, and $\partial^2 L/\partial\beta\partial\sigma=0$, as a consequence of the manner in which $\bar{x}$ was introduced into equation (2). We have written n^* to denote the number of observations which resulted in measurements of x. In the truncated and censored samples considered here, $n^*=n$, but in the selected samples, $n^*=N$. The ϕ_{ij} depend on the restrictions imposed with respect to observation of x and are available from results given in [5]. For singly truncated and singly censored samples, they are as follows:

$$
\begin{array}{lll}
 & \textit{Truncated Samples} & \textit{Censored Samples} \\
 & \phi_{11} = 1 - Z(\xi)[Z(\xi) - \xi], & \phi_{11} = 1 + Z(\xi)[Z(-\xi) + \xi], \\
(20) & \phi_{12} = Z(\xi)\{1 - \xi[Z(\xi) - \xi]\}, & \phi_{12} = Z(\xi)\{1 + \xi[Z(-\xi) + \xi]\}, \\
 & \phi_{22} = 2 + \xi\phi_{12}, & \phi_{22} = \; = 2 + \xi\phi_{12}.
\end{array}
$$

In (19) and subsequently, $\bar{\sigma}_x^2 = E(\bar{s}_x^2)$ For singly truncated and singly censored samples,

$$
(21) \qquad \bar{\sigma}_x^2 = \sigma_x^2[1 - Z(Z - \xi)].
$$

Now inverting the information matrix whose non-zero elements are given by (19), we find the non-zero elements of the asymptotic variance-covariance matrix to be

$$
(22) \quad
\begin{aligned}
& V(\hat{\alpha}) \sim \sigma^2/E(n), \quad V(\hat{\beta}) \sim \sigma^2/\bar{\sigma}^2E(n), \quad V(\hat{\sigma}) \sim \sigma^2/2E(n), \\
& V(\widehat{m}_x) \sim [\sigma_x^2/E(n^*)][\phi_{22}/(\phi_{11}\phi_{22} - \phi_{12}^2)], \\
& V(\hat{\sigma}_x) \sim [\sigma_x^2/E(n^*)][\phi_{11}/(\phi_{11}\phi_{22} - \phi_{12}^2)], \\
& \mathrm{Cov}\,(\widehat{m}_x, \hat{\sigma}_x) \sim [\hat{\sigma}_x^2/E(n^*)][-\phi_{12}/(\phi_{11}\phi_{22} - \phi_{12}^2)]
\end{aligned}
$$

In the singly restricted cases, $E(n^*) = E(n) = NI_0(\xi)$, and in the doubly restricted cases, $E(n^*) = E(n) = N[I_0(\xi_1) - I_0(\xi_2)]$ When the sampling procedure is such that $n^* = n$ is fixed, then $E(n^*) = E(n) = n$, and (22) is also applicable in that case

For a *selected sample*, $\phi_{11} = 1$, $\phi_{12} = 0$, and $\phi_{22} = 2$, since each of the N sample speciments is observed and measured without restriction with respect to x. Thus with $E(n^*) = N$, it follows from (22), that for a sample of this type

$$
(23) \quad V(\widehat{m}_x) = \sigma_x^2/N, \quad \mathrm{Cov}\,(\widehat{m}_x, \hat{\sigma}_x) = 0, \quad V(\hat{\sigma}_x) = \sigma_x^2/2N.
$$

4. AN ILLUSTRATIVE EXAMPLE

To illustrate the practical application of results obtained in this paper, we employ a sample consisting of entrance examination scores, x, and subsequent course averages, y, achieved by a group of 529 college students[1] For purposes of this illustration, we assume the minimum qualifying score on the entrance examination to be 159 5 Under this requirement, there are 517 candidates which we consider as being accepted for admission. By making appropriate further assumptions, we use the same basic data to illustrate estimation from selected, truncated and censored samples For a comparison, estimates are also com-

[1] Given by Goedicke, "Introduction to the Theory of Statistics," New York Harper and Brothers, 1953, 175–77

puted from the full (unrestricted) sample. Except for the truncated sample, we consider N as being fixed and n as a random variable. By the definition of a truncated sample, the total sample size, N, is unknown, and in that case, we consider n as fixed. Basic data for the sample selected is summarized as follows: $N=529$, $n=517$, $x_0=159.5$, $\bar{x}=164.5919$, $\bar{x}_N=164.441$, $s_x=2.847$, $\bar{s}_x=2.7036$, $\bar{y}=77.5048$, $\bar{y}_N=77.380$, $s_y=5\,405$, $\bar{s}_y=5.3470$, $\bar{r}=0\,3892$, $r=0.409$, where $\bar{x}_N$, $\bar{y}_N$, s_x, s_y, and r are based on the full sample with $N=529$, while $\bar{x}$, $\bar{y}$, $\bar{s}_x$, $\bar{s}_y$, and $\bar{r}$ are based only on observations of the $n=517$ accepted candidates. From these data, estimates were computed for the different samples, using (7), (8), (12), and (16) as applicable. Estimates of the elements of the variance-covariance matrix were formed by using the maximum likelihood estimates of the parameters in (22). Results of these computations are summarized in the accompanying table. For the full sample, since $n=N$, estimates $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}$ were computed from (7) with $\bar{y}$, $\bar{s}_x$, $\bar{s}_y$, and $\bar{r}$ replaced by $\bar{y}_N$, r, s_x, and s_y, while $\hat{m}_x$ and $\hat{\sigma}_x$ were computed from (16). Variances of the full sample estimates were computed using (22) with $\bar{\sigma}_x^2=s_x^2$, $\phi_{11}=1$, $\phi_{12}=0$, $\phi_{22}=2$, and $E(n^*)=E(n)=N$.

TABLE OF ESTIMATES AND THEIR VARIANCES

Parameters	Type of Sample			
	Complete	Selected	Censored	Truncated
m_x	164.441	164 441	164.451	164 191
σ_x	2.847	2.847	2 834	3.058
α	77.380	77.5048	77.5048	77.5048
β	0 7765	0.7697	0 7697	0.7697
σ	4 9307	4.9254	4 9254	4.9254
m_y	77 380	77.389	77.396	77.196
σ_y	5.405	5.391	5 387	5.459
ρ	0 409	0.406	0.405	0.431
ξ		−1 736	−1.747	−1.534
$E(n)$		507 162	507.671	517 (fixed)
Variance				
$V(\hat{m}_x)$	.0153	.0153	0153	.0295
$V(\hat{\sigma}_x)$	.0077	.0077	.0081	.0180
$Cov(\hat{m}_x, \hat{\sigma}_x)$	0	0	−.0002	−.0107
$V(\hat{\alpha})$	.0460	.0478	.0478	.0469
$V(\hat{\beta})$	.0057	.0071	.0071	.0064
$V(\hat{\sigma})$	.0230	0239	.0239	.0235

Considering the degree of approximation involved in their calculation, the tabulated variances reflect the varying amounts of information provided in the different types of samples, minimum information being contained in the truncated sample, and maximum information in the complete sample. The calculated variances are approximate not only due to using asymptotic values, but also due to their dependence on sample values. In comparing the truncated sample variances with those for the other samples, allowance should be made for the effective differences in sample sizes. The truncated sample with $n=517$ (fixed) corresponds to a censored or selected sample of total sample size, $N=551$.

REFERENCES

[1] Aitken, A. C., "Note on selection from a multivariate normal population," *Proceedings of the Edinburgh Mathematical Society*, 4 (1934), 106–10.

[2] Birnbaum, Z. W., "Effect of linear truncation on a multi-normal population," *Annals of Mathematical Statistics*, 21 (1950), 272–79.

[3] Birnbaum, Z. W., Paulson, E., and Andrews, F. C., "On the effect of selection performed on some coordinates of a multi-dimensional population," *Psychometrika*, 15 (1950), 191–204.

[4] Campbell, Francis L., "A study of truncated bivariate normal distributions," Doctoral Dissertation, University of Michigan, June 1945

[5] Cohen, A. C., Jr., "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples," *Annals of Mathematical Statistics*, 21 (1950), 557–69

[6] Cohen, A. C., Jr., "Estimation in truncated bivariate normal distributions," Department of Mathematics, University of Georgia, Technical Report No 2 Contract DA-01-009-ORD-288 (June 1953).

[7] Cohen, A. C., Jr., and Woodward, John, "Tables of Pearson-Lee-Fisher functions of singly truncated normal distributions," *Biometrics*, 9 (1953), 489–97.

[8] Lawly, D. N., "A note on Karl Pearson's selection formulae," *Proceedings of the Royal Society, Edinburgh, A.*, 66 (1943), 28–30.

[9] Pearson, Karl, "On the influence of natural selection on the variability and correlation of organs," *Philosophical Transactions of the Royal Society, London, A*, 200 (1903), 1–66

[10] Raj, Des, "On estimating the parameters of bivariate normal populations from doubly and singly, linearly truncated samples," *Sankhya*, 12 (1953), 277–90.

[11] Wilks, S. S., "On estimates from fragmentary data," *Annals of Mathematical Statistics*, 3 (1932), 163–96.

COMPARISON OF SOME NON-PARAMETRIC TESTS AGAINST NORMAL ALTERNATIVES WITH AN APPLICATION TO LIFE TESTING*

BENJAMIN EPSTEIN
Wayne University

1. INTRODUCTION

CONSIDER two normal populations $N(\mu_1,\sigma^2)$ and $N(\mu_2,\sigma^2)$ with common variance σ^2 (which can be assumed equal to one for convenience) and with means μ_1 and μ_2 respectively. In this paper we test the composite hypothesis $H_0:\mu_1=\mu_2$ against the composite alternative $H_1:\mu_1\neq\mu_2$ on the basis of samples of size ten drawn from each population. The performance and relative merits of four non-parametric test procedures are studied experimentally There is a close connection between this paper and recent work by Dixon and Teichroew [1] There is, however, an important difference due to the fact that we assume that the two samples in question are placed on life test, so that information about failures becomes available in an ordered way.[1] Two of the non-parametric tests considered, the rank sum and run tests, require essentially that we have information about the times to failure of all items in both samples before we can reach a decision The other two non-parametric tests, the exceedance and truncated maximum deviation tests, make essential use of the (time) ordered way in which failure data become available and thus make it possible to reach a decision long before all items fail. The experimental sampling carried out in the course of this work gives a good idea of the comparative power of the four test procedures and indicates how much one can expect to save on the average in the number of items failed in the course of reaching a decision

2. TESTS CONSIDERED IN THIS PAPER

The current study is based on 200 experiments where each experiment consists of (a) drawing a sample of ten items at random from each of two standardized normal populations, (b) placing the two samples on life test, (c) applying each of four rules of action considered below. In order to make the various rules of action comparable, we have, when

* Research sponsored by the Office of Ordnance Research of the U S Army under Contract No. DA-20-018-ORD-13272

The author wishes to thank C. K Tsao for his comments on this paper.

[1] For example, failures might be ordered in time We find it convenient in this paper to think of the observations as being ordered in this way

necessary, used randomized rules of action so as to make the Type I error the same for all test criteria. The particular Type I error chosen in this paper was $\alpha = 05$. The difference between the two populations is measured by the dimensionless parameter $d = |\mu_1 - \mu_2|/\sigma$.

Let $x_1 < x_2 < \cdots < x_{10}$ be one such ordered sample and let $y_1 < y_2 < \cdots < y_{10}$ be another such ordered sample Let $z_1 < z_2 \cdots < z_{20}$ be the ordering of the combined sample of twenty Then the rules of action for the four criteria studied here are:

(a) *The rank sum criterion S:*

Reject H_0 if the smaller (denoted as S) of the two rank sums is less than or equal to 79. H_0 is accepted otherwise See Wilcoxon [8].

(b) *The run test criterion R:*

Reject H_0 if the total number of runs (denoted as R) is less than or equal to 6. Accept H_0 if R is greater than or equal to 8 When $R=7$, perform a Bernoulli trial which accepts H_0 with probability 04 (and rejects H_0 with probability .96). See Swed and Eisenhart [6].

(c) *The exceedance criterion E_r:*

Let x_r and y_r be respectively the rth smallest observations in each of the two samples and let $w_r = \max(x_r, y_r)$ If $w_r = x_r$ count the number of y's which are $\geqq x_r$; if $w_r = y_r$ count the number of x's which are $\geqq y_r$. Denoting this number of exceedances as E_r, the test procedure is as follows: Reject H_0 if $E_r \leqq n_r - 1$, accept H_0 if $E_r \geqq n_r + 1$. When $E_r = n_r$ perform a single Bernoulli trial which rejects H_0 with probability p_r. In the present study attention was limited to the cases were $r = 1, 2, 3$. The appropriate n_r and p_r are given in Table 1.

TABLE 1

r	n_r	p_r
1	6	.32
2	4	.81
3	3	58

The theory underlying test procedures of the exceedance type is given in [3].

It may be useful to restate the exceedance rule in the language of life testing. If, for example, $r = 1$, the rule is

(i) The life test is terminated by time z_5 at the latest (z_5 is the time of occurrence of the fifth failure in both samples combined).

(ii) As soon as at least one failure in each population is observed, terminate the life test with acceptance of H_0 *provided* that this event occurs at time z_4 or earlier (z_2 or z_3).

(iii) If the first four failures all come from one of the two populations and if the fifth failure comes from the other one of the two populations, stop the test at z_5 Perform a Bernoulli trial which has only two possible outcomes A and $\bar{A}$ with $Pr(A)=32$ and $Pr(\bar{A})=.68$. If a trial is made and A occurs, reject H_0; if $\bar{A}$ occurs, accept H_0.

(iv) If the first 5 failures all come from the same population, stop testing and reject H_0.

(d) *The maximum deviation criterion* M_r:

This is a truncated maximum deviation test [7] with the truncation taking place at a time not later than $u_r=\max(x_r,y_r)$; r is decided upon in advance. For a given r, the test procedure reads as follows: Keep track of M_r, the absolute difference between the number of x failures and the number of y failures and, (i) if at any time up to and including $u_r=\max(x_r,y_r), M_r=m_r+1$ stop experimentation with the rejection of H_0. Only the decision to reject can actually be made before time u_r; (ii) if the test is actually carried to time u_r and if $M_r=m_r$ at least once (and is otherwise $\leq m_r-1$), then perform a Bernoulli trial which rejects H_0 with probability p_r; (iii) if at all times up to and including u_r, $M_r \leq m_r-1$, accept H_0 The values of M_r and p_r for $r=1$, 3, 6, 10 are given in Table 2.

TABLE 2

r	m_r	p_r
1	4	.32
3	5	.17
6	6	.95
10	6	.95

If $r=1$, we have the exceedance case (with $r=1$). If $r=10$, we have the untruncated maximum deviation test of the type considered by Smirnov [5] and Massey [4].

It seems useful to illustrate the four test criteria by means of an example. Let the first sample ($x_1<x_2< \cdots <x_{10}$) be $(-.79, -.42,$

$-.39$, $-.29$, $-.04$, $.08$, $.13$, $.22$, $.92$, 1.53) and let the second sample ($y_1 < y_2 < \cdots < y_{10}$) be ($-1.10$, $-.89$, $-.55$, $-.34$, $-.31$, $-.26$, $.35$, $.75$, 1.56, 1.91). The rearrangement of the combined sample into $z_1 < z_2 < \cdots < z_{20}$ is then (-1.10, $-.89$, $-.79$, $-.55$, $-.42$, $-.39$, $-.34$, $-.31$, $-.29$, $-.26$, $-.04$, $.08$, $.13$, $.22$, $.35$, $.75$, $.92$, 1.53, 1.56, 1.91). Criteria (a), (b), (c), and (d) work out as follows on this example:

(a) In the combined sample of 20, the sum of the ranks of the x's is 108 and the sum of the ranks of the y's is 102. Thus H_0 is accepted since 102, the smaller of the two rank sums exceeds 79.

(b) To apply the run test we write the combined ordered sample as *yyxyxxyyxyxxxxyyxxyy* The total number of runs equals 11 and so H_0 is accepted.

(c) Exceedance criteria $E_r(r=1, 2, 3)$ all lead to the acceptance of H_0 after 3, 5, and 6 observations respectively.

(d) The maximum deviation criteria M_r $(r=1, 3, 6, 10)$ all lead to the acceptance of H_0 after 3, 6, 12, and 20 observations respectively.

3. COMPARISON OF THE POWER OF THE FOUR TEST CRITERIA

In the following table we summarize the experimental findings for the 200 pairs of samples, where each sample is of size ten. These samples were drawn from Wold's tables of random normal deviates [9]. Samples corresponding to the cases where $d = \left|(\mu_1-\mu_2)/\sigma\right| = 1, 2, 3$ were ob-

TABLE 3

OBSERVED PROBABILITY OF ACCEPTING $H_0(d=0)$ BASED ON 200 PAIRS OF SAMPLES, EACH OF SIZE TEN

$d=\left\|\frac{\mu_1-\mu_2}{\sigma}\right\|$	Rank Sum	Run	Exceedance			Maximum Deviation		
			$r=1$	$r=2$	$r=3$	$r=3$	$r=6$	$r=10$
0	935	.965	.95	.96	.96	.955	.945	.945
1	.485	795	.655	.65	.60	.575	.555	.555
2	.015	.275	.16	.12	.10	.065	.045	.045
3	0	.02	025	0	0	0	0	0

tained by adding 1, 2, and 3 to the x_i's and leaving the y_i's unchanged. This was done for convenience. It might have been better to have used 200 different pairs of samples for each of the four values of d. The author doubts, however, that this would make any appreciable change in the overall pattern presented by Table 3.

The following remarks are pertinent:

(i) As r increases, there appears to be a slight improvement in the power of exceedance and maximum deviation tests. It happens that the truncated maximum deviation test for $r=6$ has the same experimental O C. curve as the untruncated maximum deviation test for the particular samples being reported on in this paper

(ii) The maximum deviation test has slightly better power than the exceedance test for the particular samples being reported on in this paper

(iii) Ranked in order of power we have: Rank sum, best; run test, worst; exceedance and maximum deviation tests in between.

In order to be able to make more positive and more general statements, particularly in (i) and (ii), we would need much more in the way of experimental evidence. To settle the question completely awaits the theoretical treatment of what seems to be a complicated analytical problem.

4 AVERAGE SAMPLE SIZES

It is of interest to report how many observations were required on the average to reach a decision if one adopts a decision rule E_r or a decision rule M_r. In Table 4 we give data on rules E_1, E_2, and E_3 and in Table 5 we give data on M_3.

TABLE 4

AVERAGE NUMBER OF ITEMS FAILED IN REACHING A DECISION IF EXCEEDANCE RULE E_r IS USED. THE AVERAGES GIVEN ARE BASED ON WHAT WAS OBSERVED IN 200 EXPERIMENTS, EACH EXPERIMENT CONSISTING OF A PAIR OF SAMPLES EACH OF SIZE TEN

$d=\left\|\frac{\mu_1-\mu_2}{\sigma}\right\|$	E_1	E_2	E_3
0	2 85	4.99	7.25
1	3 84	6 42	8.64
2	4 77	7 29	8.85
3	4.96	7.07	8 17

TABLE 5

AVERAGE NUMBER OF ITEMS FAILED IN REACHING A DECISION IF TRUNCATED MAXIMUM DEVIATION RULE M_t IS USED

$d = \left\lvert \frac{\mu_1 - \mu_2}{\sigma} \right\rvert$	Average Number (Based On 200 Experiments)
0	7 17
1	8.02
2	6.80
3	6.11

Tables 4 and 5 indicate clearly why exceedance and truncated maximum deviation procedures should be given serious consideration, if data become available in an ordered way, as they do in life testing. One may well be willing to sacrifice some power, if this means that one can in this way attain a substantial reduction in the number of items failed in the course of reaching a decision Of course one should bear in mind that while the tests considered here will be reasonably effective in detecting whether the two distributions $f(x)$ and $g(y)$ differ in location, they may be quite insensitive to other important differences This is an inherent difficulty in all non-parametric procedures and this is why the choice of a non-parametric procedure should be based at least in part on what we know about the underlying distributions in an a priori way or from data collected in the past.

5 CONCLUSION

We have presented the results of a sampling experiment of moderate size to indicate the possible usefulness of using exceedance or truncated maximum deviation tests when the data from each of two samples become available in an ordered way. The underlying distributions are assumed to be normal While the experimental results in this paper are reported only for the case where the common sample size is ten, it is safe to conjecture that similar results would be found for other sample sizes

BIBLIOGRAPHY

1. Dixon, W. J , and Teichroew, D., "Some sampling results on the power of nonparametric tests against normal alternatives," Abstract, *Annals of Mathematical Statistics*, 25 (1954), 175

2. Dixon, W. J., "Power under normality of several nonparametric tests," *Annals of Mathematical Statistics*, 25 (1954), 610–14.
3. Epstein, B., "Tables for the distribution of the number of exceedances," *Annals of Mathematical Statistics*, 25 (1954), 762–8.
4. Massey, F. J., Jr., "The distribution of the maximum deviation between two cumulative step functions," *Annals of Mathematical Statistics*, 22 (1951), 125–8.
5. Smirnov, N., "Tables for estimating the goodness of fit of empirical distributions," *Annals of Mathematical Statistics*, 19 (1948), 279–81
6. Swed, F. S., and Eisenhart, C., "Tables for testing randomness of grouping in a sequence of alternatives," *Annals of Mathematical Statistics*, 14 (1943), 66–87
7. Tsao, C. K., "An extension of Massey's distribution of the maximum deviation of two cumulative step functions," *Annals of Mathematical Statistics*, 25 (1954), 587–92.
8. Wilcoxon, F., "Probability tables for individual comparisons by ranking methods," *Biometrics*, 3 (1947), 119–22.
9. Wold, H., "Random normal deviates," *Tracts for Computers No. 25*, Cambridge University Press (1948).

ON THE DISTRIBUTION OF A POSITIVE RANDOM VARIABLE HAVING A DISCRETE PROBABILITY MASS AT THE ORIGIN*

JOHN AITCHISON
University of Cambridge

In a number of situations we are faced with the problem of determining efficient estimates of the mean and variance of a distribution specified by (i) a non-zero probability that the variable assumes a zero value, together with (ii) a conditional distribution for the positive values of the variable. This estimation problem is analyzed and its implications for the Pearson type III, exponential, lognormal and Poisson series conditional distributions are investigated Two simple examples are given

1 THE PROBLEM

THE nature of the problem to be discussed is best introduced by examples In a study of household expenditures it is often of interest to estimate, from a sample of household budgets, the mean expenditure per household on a certain commodity, say children's clothing. Over the period of the investigation it may well happen that a number of households in the sample spend nothing on children's clothing whereas the expenditures by the remainder of the households necessarily arise from the distribution of a positive variable, probably skew and possibly approximated by a lognormal curve. If such is the case, then clearly the correct procedure in any analysis is to recognize explicitly this dichotomy of the population into the categories, spender and non-spender. This type of situation is not, however, confined to the case of a continuous variable; it occurs also for discrete variables For example, in a household composition study we may wish to investigate the distribution of the number of children in a household. This distribution is sometimes Poisson except that the number of households with no children is considerably larger than is suggested by Poisson theory. Again one solution of the difficulty is to assume that there is a proportion of households containing no children while the remainder is distributed as a truncated Poisson distribution.

Such problems lead us to consider a random variable X with the following properties There is a non-zero probability θ that X is zero

* This paper is a development of some of the estimation problems discussed by Utting and Cole [5]. The author wishes to express his indebtedness to J A C Brown of the Department of Applied Economics for helpful criticism and for suggesting the application of the Poisson series distribution to the analysis of household composition.

and hence a probability $1-\theta$, that X is non-zero; further the distribution of X conditional on $X\neq O$ is some well-known distribution of a positive variable, either continuous or discrete. This we may write:

$$P\{X = 0\} = \theta, \tag{1}$$

$$P\{X > 0\} = 1 - \theta, \tag{2}$$

and, for the continuous case

$$P\{X \subset (x, x + dx) | x > 0\} = g(x)dx, \tag{3}$$

where $g(x)$ is the conditional frequency function; and so

$$P\{X \subset (x, x + dx)\} = (1 - \theta)g(x)dx, x > 0 \tag{4}$$

If α and β are the mean and variance respectively of the $g(x)$ distribution and γ and δ are the corresponding parameters of X then

$$\gamma = (1 - \theta)\alpha \tag{5}$$

and

$$\delta = (1 - \theta)\beta + \theta(1 - \theta)\alpha^2 \tag{6}$$

We discuss in this paper the problem of efficient estimation of γ and δ.

2 EFFICIENT ESTIMATION

In this section we state some general results proved in the Appendix which allow us, under certain circumstances, to obtain best unbiased estimators of γ and δ; by the term *best unbiased estimator* we mean an unbiased estimator having minimum attainable variance (see, for example, Rao [3]) Let us suppose that, for the purpose of estimation, we have available a random sample S of size n from the population and that r of the sample values are zero while the remaining $(n-r)$ are $x_1, x_2, \cdots, x_{n-r}$ Then the following results hold

(i) If, for a sample of size m from the $g(x)$ population a sufficient unbiased estimator of α, say $a_{(m)}$, exists then

$$\begin{aligned} c &= \left(1 - \frac{r}{n}\right)a_{(n-r)},\ r < n, \\ &= 0,\ r = n, \end{aligned} \tag{7}$$

is a best unbiased estimator of γ

The twofold definition of c is necessary since $a_{(n-r)}$ is not defined for

$r=n$. If $a_{(m)}$ is the arithmetic mean of m sample values then c becomes the mean of the sample S (including zero values), namely

$$c = \frac{1}{n} \sum_{i=1}^{n-r} x_i \tag{8}$$

and the variance of the estimator in this case is

$$\operatorname{var} \{c\} = \frac{\delta}{n} \tag{9}$$

A result similar to (i) holds for δ provided that jointly sufficient estimators of α (and hence of α^2) and β exist.

(ii) If $e_{(m)}$ and $f_{(m)}$ are jointly sufficient unbiased estimtaors of α^2 and β respectively for a sample of size m, then

$$d = \left(1 - \frac{r}{n}\right) f_{(n-r)} + \frac{r}{n}\left(1 - \frac{r-1}{n-1}\right) e_{(n-r)}, \ r < n,$$
$$= 0, \ r = n, \tag{10}$$

is a best unbiased estimator of δ

It is seldom that such jointly efficient estimators of α and β occur Often, however, β depends on α so that $a_{(m)}$ is sufficient for both α and β; an important case is $\beta = K\alpha^2$, for which we have the following property

(iii) If $\beta = K\alpha^2$ and $a_{(m)}$, the sufficient unbiased estimator of α, is the sample mean then

$$\delta = (1 - \theta)(K + \theta)\alpha^2 \tag{11}$$

and

$$d = \left\{K + (1-K) \ \frac{r}{n} - \frac{r(r-1)}{n(n-1)}\right\} a_{(n-r)}{}^2 \Big/ \left\{1 + \frac{K}{n-r}\right\}, r < n,$$
$$= 0, \ r = n, \tag{12}$$

is a best unbiased estimator of δ

3 APPLICATION TO PARTICULAR DISTRIBUTIONS WITH EXAMPLES

It is interesting to apply the estimator procedure of the preceding section to a number of particular conditional distributions

(i) ***Pearson type III distribution***

For the Pearson type III distribution

$$g(x) = \left(\frac{p}{\alpha}\right)^p \frac{x^{p-1}e^{-px/\alpha}}{\Gamma(p)}, \quad x > 0, \tag{13}$$

where we assume p to be known. For the mean α of this distribution the sample mean is a sufficient unbiased estimator so that

$$c = \frac{1}{n}\sum_{i=1}^{n-r} x_i \tag{14}$$

is a best unbiased estimator of $\gamma=(1-\theta)\alpha$. Here $\beta=\alpha^2/p$ and so

$$\delta = (1-\theta)\left(\frac{1}{p}+\theta\right)\alpha^2, \tag{15}$$

and a best unbiased estimator of δ is given by (12) with $K=1/p$.

(ii) *Exponential distribution*

For the exponential distribution,

$$g(x) = \frac{1}{\alpha}e^{-x/\alpha}, \; x > 0, \tag{16}$$

and this is the special case $p=1$ of the Pearson type III distribution so that no new theory arises The estimator of δ, however, simplifies to

$$\begin{aligned} d &= \frac{1+\dfrac{r-1}{n}}{1-\dfrac{r-1}{n}} \frac{\left\{\sum_{i=1}^{n-r} x_i\right\}^2}{n(n-1)}, \; r < n, \\ &= 0, \; r = n. \end{aligned} \tag{17}$$

(iii) *Lognormal distribution*

If the conditional distribution is lognormal with parameters μ and σ^2 then

$$g(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}(\log x - \mu)^2\right\}, \; x > 0 \tag{18}$$

Here

$$\alpha = e^{\mu+\frac{1}{2}\sigma^2}, \tag{19}$$

so that

$$\beta = e^{2\mu+\sigma^2}(e^{\sigma^2}-1), \tag{20}$$

so that

$$\gamma = (1-\theta)e^{\mu+\frac{1}{2}\sigma^2}, \tag{21}$$

and

$$\delta = (1-\theta)\cdot e^{2\mu+\sigma^2}\left\{e^{\sigma^2}-(1-\theta)\right\}. \tag{22}$$

Finney [2] has obtained by considerations similar to those of this paper best unbiased estimators of α and β, and as an extension of his theory it may be shown that

$$
\begin{aligned}
c &= \left(1 - \frac{r}{n}\right) e^{\bar{y}} \psi_{(n-r)}(\tfrac{1}{2}s^2), \ r < n - 1, \\
&= \frac{x_1}{n}, \ r = n - 1, \\
&= 0, \ r = n,
\end{aligned}
\tag{23}
$$

and

$$
\begin{aligned}
d &= \left(1 - \frac{r}{n}\right) e^{2\bar{y}} \left\{\psi_{(n-r)}(2s^2) \right. \\
&\qquad \left. - \left(1 - \frac{r}{n-1}\right) \psi_{(n-r)} \left(\frac{n-r-2}{n-r-1} s^2\right)\right\}, \ r < n - 1, \\
&= \frac{x_1^2}{n}, \ r = n - 1, \\
&= 0, \ r = n,
\end{aligned}
\tag{24}
$$

are best unbiased estimators of γ and δ respectively, where

$$
y_i = \log x_i, \ i = 1, 2, \cdots, n - r, \tag{25}
$$

$$
\bar{y} = \frac{1}{n - r} \sum_{i=1}^{n-r} y_i, \ r < n - 1, \tag{26}
$$

$$
s^2 = \frac{1}{n - r - 1} \sum_{i=1}^{n-r} (y_i - \bar{y})^2, \ r < n - 1, \tag{27}
$$

and[1]

$$
\begin{aligned}
\psi_m(t) = 1 + \frac{m-1}{m} t &+ \frac{(m-1)^3}{2!m^2(m+1)} t^2 \\
&+ \frac{(m-1)^5}{3!m^3(m+1)(m+3)} t^3 + \cdots .
\end{aligned}
\tag{28}
$$

It can be shown that $1/n \sum_{i=1}^{n-r} x_i$ is less efficient than c in this case.

Example.[2] In a household expenditure inquiry carried out by the Ministry of Food in 1950, a sample of 1143 British households was

---

[1] A table of values of the function $\psi_m(t)$ will be published in a forthcoming monograph on the lognormal distribution by the Department of Applied Economics, University of Cambridge.

[2] The data for this and the following example have been obtained from The National Food Survey by courtesy of the Ministry of Food.

taken; one of the items in the classification gives expenditures in pence per household on sweet biscuits. Of the 1143 households only 512 bought this commodity and their expenditures appear to come from a lognormal population. The relevant sample data are:

$$n = 1143 \qquad \bar{y} = 3.664$$
$$r = 631 \qquad s^2 = 0.3721$$

The estimator c of γ obtained from (23) is then

$$\begin{aligned} c &= \left(1 - \frac{631}{1143}\right) e^{3.664} \psi_{(512)}(0.1860) \\ &= 20.95, \end{aligned}$$

as compared with the value of 20 25 given by the ordinary sample mean. The value of d in this case is 4481 so that the estimated standard error of the sample mean is 1 97 and the standard error of c is necessarily less than this

(iv) *Truncated Poisson distribution*

The truncated Poisson distribution[3] provides an application of the theory to a discrete variable. In this case

$$\begin{aligned} g(x) &= P\{X = x \mid x > 0\} \\ &= \frac{e^{-\mu}}{1 - e^{-\mu}} \frac{\mu^x}{x!}, \quad x = 1, 2, \cdots \end{aligned} \tag{29}$$

and

$$\alpha = \frac{\mu}{1 - e^{-\mu}} \cdot \tag{30}$$

The sample mean is again a sufficient unbiased estimator[4] of α and so c as given by (8) is a best unbiased estimator of γ where

$$\gamma = \frac{(1 - \theta)\mu}{1 - e^{-\mu}} \cdot \tag{31}$$

It does not seem possible to find a simple expression for the estimator d in this case.

[3] See, for example, David and Johnson [1]

[4] See Tukey [4]

Example. The data of Table 1 analyze by number of children (under fourteen years of age) per household a sample of 4021 British households in 1950.

TABLE 1

NUMBER OF HOUSEHOLDS CONTAINING GIVEN NUMBER OF CHILDREN

No. of Children	0	1	2	3	4	5	6	7	8	9
(i) Observed	2303	831	565	212	67	23	15	3	1	1
(ii) Poisson	1856	1435	554	143	28	4	1	—	—	—
(iii) Truncated Poisson	2303	822	546	242	81	21	5	1	—	—

Rows (ii) and (iii) of the table show the results of fitting a complete Poisson distribution (estimated $\mu = 0.773$) and a truncated Poisson distribution (estimated $\mu = 1.33$) as described by (29) The complete Poisson distribution clearly does not give an adequate fit due to the extra large proportion of households with no children; the truncated distribution gives a better approximation and a best unbiased estimate of the mean number of children per household is simply the sample mean 0.773.

4 FURTHER CONSIDERATIONS

The limitations that the discrete probability mass is at the origin rather than at some other point, and that the conditional variable is essentially positive, may be removed without unduly complicating the theory, we have not thought it worth-while to develop this extension because of the lack of any obvious practical application. It would also have been interesting to compare the efficiency of other possible estimators with the estimators derived for the special distributions but this particular problem has also been left aside

APPENDIX: DERIVATION OF THEORETICAL RESULTS

In this Appendix we derive the theoretical results of Section 2. The essential idea underlying the proofs is a property of jointly sufficient estimators: any function of jointly sufficient estimators is a best unbiased estimator of its expectation (cf. Rao [3, p 149])

As in Section 2 the random sample S consists of r zero values and $(n-r)$ other values $x_1, \cdots, x_{n-r}$ For the $g(x)$ population, $a_{(m)}$ is a sufficient unbiased

estimator of α for a sample of m values. If the distribution of X depends on parameters $\lambda, \cdots$ in addition to θ and α, then the likelihood function L of the sample may be written in the form

$$L(S \mid \theta, \alpha, \lambda, \cdots) = \binom{n}{r} \theta^r (1-\theta)^{n-r} h(a_{(n-r)}, \alpha, \lambda, \cdots) k(S \mid \lambda, \cdots)$$

where h is a frequency function containing the sample values only in the form $a_{(n-r)}$, and k is a frequency function independent of θ and α Hence

$$L = L_1\left(\frac{r}{n}, a_{(n-r)}, \theta, \alpha, \lambda \cdots\right) L_2(S \mid \lambda, \cdots)$$

where L_1 is a frequency function containing the sample values only in the forms r/n and $a_{(n-r)}$, and L_2 is a frequency function independent of θ and α. Thus r/n and $a_{(n-r)}$ are jointly sufficient estimators of θ and α Consider the sample function:

$$\begin{aligned} c &= \left(1 - \frac{r}{n}\right) a_{(n-r)}, \ r < n \\ &= 0, \ r = n \end{aligned}$$

The expectation of the estimator is given by

$$\begin{aligned} E\{c\} &= P\{r = n\} E|_{r=n}\{c\} + P\{r < n\} E|_{r<n}\{c\} \\ &= P\{r < n\} E_{r|r<n}\left\{\left(1 - \frac{r}{n}\right) E|_{r=\text{const}<n}[a_{(n-r)}]\right\} \\ &= P\{r < n\} E_{r|r<n}\left\{\left(1 - \frac{r}{n}\right)\alpha\right\} \\ &= \alpha E\left(1 - \frac{r}{n}\right) \\ &= (1-\theta)\alpha \\ &= \gamma \end{aligned}$$

so that, from the property of sufficient estimators referred to above, c is a best unbiased estimator of γ.

The proofs of the results (ii) and (iii) of Section 2 proceed in exactly the same manner and their details need not be reproduced here

REFERENCES

[1] David, F. N., and Johnson, N. L., "The truncated Poisson," *Biometrics*, 8 (1952), 275–85.

[2] Finney, D. J., "On the distribution of a variate whose logarithm is normally distributed," *Journal of the Royal Statistical Society*, Series B, 7 (1941), 155–61.

[3] Rao, C. R., *Advanced Statistical Methods in Biometric Research* New York: John Wiley and Sons, Inc., 1952

[4] Tukey, John W., "Sufficiency, truncation and selection," *Annals of Mathematical Statistics*, 20 (1949), 309–11.

[5] Utting, J. E. G., and Cole, Dorothy, "Sample surveys for the social accounts of the household sector," *Bulletin of the Oxford Institute of Statistics*, 15 (1953), 1–24.

INCREASING THE EFFECTIVE LENGTH OF SHORT TIME-SERIES FOR THE PURPOSE OF ESTIMATING AUTOREGRESSIVE PARAMETERS*

ABBOTT S WEINSTEIN
New York State Department of Commerce

Techniques of overcoming biases in serial correlation coefficients computed from short linear autoregressive time-series are considered Two alternative estimates of autocorrelation coefficients are suggested as probable improvements over the usual estimates based on averages of observations made over a seasonal period. The alternative estimates are obtained, for example, from 12 yearly series, one for each month, and are:

1. an average of serial coefficients computed from each yearly series,
2 pooled sums of squares and lagged products for each series formed into a single serial correlation coefficient.

It is well known that the application of conventional least squares regression analysis to economic time-series is complicated by (1) autocorrelated error terms, (2) errors of observation, and (3) simultaneous interrelationships among the variables.[1] Even in the uncomplicated case where the error term is distributed normally, the maximum likelihood solution may lead to estimators that are biased to order $1/n$, as noted by Kendall [8] Where only one of the complications is present, large-sample methods of overcoming it are available, but these are similarly biased for short series With more than one of the complications, the problem is greater and a much larger number of observations is essential.[2]

In view of the brevity of most economic series, the effect of these considerations is especially serious It is the purpose of this paper to describe methods designed to increase the effective number of observations of the series, thereby reducing the short-series limitations of available analytical techniques. The discussion deals specifically with the elimination of bias from serial correlation coefficients used in the estimation of autoregressive parameters.

A series of annual data consists either of annual averages of periodic observations, or of one observation per year made at roughly the same

* The author wishes to express his appreciation to Alfred D Basch and Ethel E. Metzendorf of the New York State Department of Commerce and Joseph Lev of the New York State Department of Civil Service for their encouragement and helpful comments.

[1] Cochrane and Orcutt [4].

[2] Orcutt and Cochrane [12].

date each year. We consider here series of annual data that can be represented usefully by the two-lag autoregressive process,[3]

$$x_{t+2} + \alpha_1 x_{t+1} + \alpha_2 x_t = v_{t+2}, \tag{1}$$

where v_t is a random disturbance with $E\{v_t\}=0$ and the t's are equidistant points in time The specific problem considered is the elimination of bias from the serial correlation coefficients used to estimate the α's

As shown by Orcutt [11], the serial correlation coefficient is biased in two respects: downward as a result of using the sample mean where the parent mean is not known, and toward zero because of the skewed distribution of the coefficients. Since the first two serial coefficients are positive in most economic series, the entire bias in these coefficients will usually be downward. In the method of estimating the α's considered below, serial correlation coefficients are used for the first two lags only

A correction for downward bias has been suggested by Quenouille [14] Based on his conclusion, drawn from an empirical study, that the downward bias of a serial correlation coefficient is inversely proportional to n, Quenouille proposed replacing r_s with

$$r_s \text{ (corrected for bias)} = 2r_s - \tfrac{1}{2}({}_1r_s + {}_2r_s) \tag{2}$$

where r_s is the serial correlation coefficient for the sth lag as computed from a given series, and ${}_1r_s$ and ${}_2r_s$ are the corresponding coefficients based on each of the two halves of the same series

The correction, intriguing because of its simplicity, loses some of its appeal when a series of, say, only 14 annual observations is involved. Using the correction would require basing ${}_1r_s$ and ${}_2r_s$ on series with $n=7$ Most economic series are internally comparable for very few years, and serial correlation coefficients estimated from such short series will be badly biased.[4] Hence, where we need it most, the correction is of the least apparent use.

What is needed, since it is practically impossible to get observations for more years, is an approach that will give us more annual observations for the years for which data are available. Where data are observed monthly, we have not only one annual series, but actually 12 such series, one for each month

In order for the methods proposed here to be valid, it is necessary that

---

[3] For a discussion of the autoregressive scheme, see Wold [23], Bartlett [2], and Kendall [6, 7, 8]

[4] Sastry [19] has shown that there is appreciable bias in certain autoregressive series of 200 observations or more

each of the 12 series have the same parametric structure. While objective tests of this assumption are not known, it does have some logical and empirical foundation We are, in fact, attempting to analyze annual serial movement by assuming a particular type of autoregressive process and then estimating the generating constants. If this approach is to have any meaning, it would seem that the annual movement should characterize the whole year, not an isolated month of the year. This is not very different from the assumption inherent in using the series of annual averages, themselves, as representative of the entire annual movement Moreover, graphic records of actual series prepared by the author tend to support the assumption of a single parametric structure. Nevertheless, it is highly advisable to examine this assumption as it applies to a particular series before applying the methods suggested here.

The model (1) requires further the assumption of stationarity—that the series have no trend In practice, however, especially in studying short series, it is often regarded as prudent to use original observations that are only approximately stationary, since the effect of trend removal is uncertain and possibly more detrimental than the existence of a mild trend. Some other models which are in some ways preferable to (1) (Orcutt's [11], for example) are, in themselves, not stationary

Where the 12 series can be treated as if they have the same parametric structure, two possibilities suggest themselves. I. Coefficients, $r_{s j}$, might be computed for each lag, s, from each of the 12 annual series, and the mean of the $12 r_{s j}$'s obtained. The mean would have the same bias as the component $r_{s j}$'s and an appropriate correction for bias would have to be applied; II The sums of squares and lagged products for each series might be pooled to produce a single serial correlation coefficient, r_s. Method II is equivalent to computing, from the single long series of monthly observations, serial correlation coefficients with a lag 12 times the annual lag We have, where the mean of the series is zero,

$$(3)\qquad \frac{\sum_{j=1}^{12}\sum_{i=1}^{n-s} x_{ij}x_{i+s,j}}{\sqrt{\left[\sum_{j=1}^{12}\sum_{i=1}^{n-s} x_{ij}^2\right]\left[\sum_{j=1}^{12}\sum_{i=1+s}^{n} x_{ij}^2\right]}} = \frac{\sum_{i=1}^{N-k} x_i x_{i+k}}{\sqrt{\sum_{i=1}^{N-k} x_i^2 \sum_{i=1+k}^{N} x_i^2}}$$

which approximates r_k and r_s, where year $i=1, 2, \cdots, n$; month $j=1, 2, \cdots, 12=$Jan., Feb , $\cdots$, Dec ; $N=12n$, $k=12s$. This is almost analogous to rearranging the series into one long *annual* series

consisting of first the n January observations, followed by the n February observations, etc. However, as is implicit in (3), the series is "segmented"; it should be clear that this is necessary, since there would be no meaning in pairing the last terms of the series of January observations with the first terms of the February series, etc. Even the longer series may be biased, and a correction is probably necessary.

In order to take advantage of Quenouille's correction in short series, the following methods are proposed:

OVERCOMING BIAS, METHOD I

An estimate of ρ_s, corrected for bias, is

$$(4)\qquad r_s \text{ (corrected for bias)} = 2\bar{r}_{s(j,j+1)} - \bar{r}_{sj}$$

where s is the number of years lagged, j is the month, and where

$$(5)\qquad \bar{r}_{s(j,j+1)} = \frac{1}{6}\sum_{j=1}^{11} r_{s(j,j+1)}, \qquad j = 1, 3, \cdots, 11,$$

and

$$(6)\qquad \bar{r}_{sj} = \frac{1}{12}\sum_{j=1}^{12} r_{sj}, \qquad j = 1, 2, \cdots, 12.$$

The basic serial correlation coefficients[5] are computed from each of the individual segments, j, of the series of variates, X_{ij},

$$(7)\qquad r_{sj} = \frac{2\sum_{i=1}^{n-s} X_{ij}X_{i+s,j} - 2(n-s)\bar{X}_j^2}{\sum_{i=1}^{n-s} X_{ij}^2 + \sum_{i=1+s}^{n} X_{ij}^2 - 2(n-s)\bar{X}_j^2}.$$

The r_{sj}'s are averaged according to (6). Combining the 12 segments in pairs to form six longer segments, the coefficient based on any one of the combined segments is

$$(8)\ r_{s(j,j+1)} = \frac{2\sum_{i=1}^{n-s} X_{ij}X_{i+s,j} + 2\sum_{i=1}^{n-s} X_{i,j+1}X_{i+s,j+1} - 4(n-s)\bar{X}_{j,j+1}^2}{\sum_{i=1}^{n-s} X_{ij}^2 + \sum_{i=1+s}^{n} X_{ij}^2 + \sum_{i=1}^{n-s} X_{i,j+1}^2 + \sum_{i=1+s}^{n} X_{i,j+1}^2 - 4(n-s)\bar{X}_{j,j+1}^2}$$

[5] In this discussion and in the illustration which follows, a modified definition of the serial correlation coefficient is used. This definition and several others are discussed by the author in another paper which is in progress.

where j is the segment, n is the number of observations in either segment, and the single mean of the two series, $\overline{X}_{j,j+1}$, is

$$\frac{1}{2n}\sum_{i=1}^{n}(X_{ij}+X_{i,j+1}).$$

Except for $\overline{X}_{j,j+1}$, all the elements of the formula are available directly from computations already made in obtaining (7). The $r_{s(j,j+1)}$'s are averaged according to (5).

OVERCOMING BIAS, METHOD II

With as many observations as we have in the 12 series, we might prefer to compute a single coefficient for each lag from the whole series since the bias is a function of the size of the sample.

Thus we might apply Quenouille's correction directly to the entire "long series":

$$r_s = 2r_s - \tfrac{1}{2}({}_1r_s + {}_2r_s) \tag{9}$$

where r_s is the serial correlation coefficient for the sth lag as computed from a given series, and ${}_1r_s$ and ${}_2r_s$ are the corresponding coefficients based on each of the two halves of the same series

As discussed above, the long series should be treated as if it were still segmented; i e , r_s based on the whole series is defined by

$$\frac{2\sum_{j=1}^{12}\sum_{i=1}^{n-s}X_{ij}X_{i+s,j}-24(n-s)\overline{X}^2}{\sum_{j=1}^{12}\sum_{i=1}^{n-s}X_{ij}^2+\sum_{j=1}^{12}\sum_{i=1+s}^{n}X_{ij}^2-24(n-s)\overline{X}^2}, \tag{10}$$

where n is the number of observations, i, in each segment, j, and $\overline{X}$ is the mean of all the variates in the series. We are limited in the gain in "length" we might expect, since using (10), "end-effects"[6] will retain the same proportion of the variation they had in (7) and (8). Where the series is segmented, the expected proportion of "end-effects" is not represented by the ratio $(n-s)/n$ which would approach unity if the series were really a single set of observations Instead, the proportion is $(n-qs)/n$ (where q is the number of segments) and remains a constant. Therefore, even the estimate based on the long series (segmented)

[6] "End-effects," absent in ordinary correlation, are found in serial correlation and arise from the fact that in the numerator of the formula for r, for example, the first and last s terms are paired only once, while the middle $n-2s$ terms are each used twice.

will be considerably biased and the application of Quenouille's correction (9) should lead to an improvement.[7]

The Methods Illustrated. We use Kendall's [7] experimental series No 1, which had been constructed by Kendall to conform to (1) with $\alpha_1 = -1.1$ and $\alpha_2 = +0.5$. His v_t was based on tables of random numbers and is distributed rectangularly from -49 to $+49$ In a series of infinite length having the same constants, $\rho_1 = +.733$ and $\rho_2 = +.306$

In order that the experimental series represent the 12 annual series or the single long segmented series discussed above, it has been broken into 12 segments ("months") of 14 observations ("years") each Series of roughly this length are likely to be encountered in practice and have received considerable attention in the literature. The first segment consists of the first 14 terms of the original series; the second, of terms 15 through 28, etc In this manner, the first 168 of Kendall's 480 terms are used The 12 segments have also been combined to form a series analogous to annual averages. We have, as terms of the "average" segment,

$$X_i = \sum_{j=1}^{12} X_{ij}.$$

Actually, the segment consists of totals rather than averages, this saves a step in the computation without changing the value of the coefficients. The coefficients computed from specified segments are shown in Table I

The coefficients based on the "average" segment are $+.6301$ for $s=1$ and $+.0493$ for $s=2$ Only one of the r_{1j}'s and two of the r_{2j}'s are higher than their respective true value, with no bias, roughly half would be higher. The bias is directly indicated by the fact that the means of of the r_{1j}'s and r_{2j}'s are each far below the known parameters, as are the r_s's based on the average segment

The annual averages have the advantage of being independent of seasonality and monthly serial correlation In this example, moreover, as should be expected, serial coefficients computed from the average segment are higher than the means of those of the 12 series. This is partly because of the smoothing effect of averaging and partly because the mean of the series of averages is a better estimator of the true mean than is that of any of the 12 segments However, where n is as small as in the examples considered here, estimates based on the series of annual averages are also excessively biased, and with so short a single

[7] Even where the series is not segmented, bias due to the mean, alone, is found in certain series with 200 observations or more, as noted above

TABLE I

SERIAL CORRELATION COEFFICIENTS COMPUTED FROM INDIVIDUAL AND GROUPED SEGMENTS OF KENDALL'S SERIES 1*

($\rho_1 = + 733$, $\rho_2 = + 306$)

Segment, j	Serial Correlation Coefficient (and formula number)							
	Method I		Method II		Method I		Method II	
	r_{1j} (7)	$r_{1(j, j+1)}$ (8)	${}_1r_1$, ${}_2r_1$ (10)†	r_1 (10)	r_{2j} (7)	$r_{2(j, j+1)}$ (8)	${}_1r_2$; ${}_2r_2$ (10)†	r_2 (10)
1	4577	6174	.7003	.6778	.0875	.3517	.3499	.2310
2	2539				−.1995			
3	2780	7015			− 0571	.3619		
4	7978				3977			
5	6426	5975			− 0329	−.0054		
6	.5567				.0232			
7	.5036	6133	.6359		− 3137	− 0277	.0420	
8	6765				1257			
9	5334	6274			− 4290	−.0616		
10	6029				− 1035			
11	6302	6803			2375	2908		
12	7060				3152			
Mean	5533	6396	6681	.6778	0043	1516	1960	2310
Variance	0243	.0014	0010	—	0563	0343	0237	—

* The coefficients based on the 'average' segment are + 6301 for $s = 1$ and + 0493 for $s = 2$

† In this case (10) is applied separately to each of the two halves of the series rather than to the whole series as discussed in the text.

series, the dependability of known corrections for the bias is uncertain

In the present example, r_1 and r_2 computed according to (10), while below the true values, are nearer than $\bar{r}_{sj}$ or $\bar{r}_{s(j,\ j+1)}$. Note the improvement in the estimates, compared to the true values of .733 and .306, as n increases

Applying Method I, we have

$$r_1 \text{ (corrected for bias)} = 2(.6396) - .5533 = + .7259$$

which is very close to the true value, +.733, and

$$r_2 \text{ (corrected for bias)} = 2(.1516) - .0043 = +.2989$$

which is similarly close to its true value, +.306.

Applying Method II, we have

$$r_1 \text{ (corrected for bias)} = 2(.6778) - .6681 = +.6875, \quad \text{and}$$

$$r_2 \text{ (corrected for bias)} = 2(.2310) - .1960 = +.2660.$$

Estimating Autoregressive Parameters A straightforward method of estimating the α's in (1) from a series of observations requires simply the computation of serial correlation coefficients for each of the first two lags (r_1 and r_2) and substituting for the unknown autocorrelation coefficients (ρ_1 and ρ_2, respectively) in the recurrence relationship,

$$\begin{aligned} \rho_1\alpha_0 + \rho_0\alpha_1 + \rho_1\alpha_2 &= 0 \\ \rho_2\alpha_0 + \rho_1\alpha_1 + \rho_0\alpha_2 &= 0 \end{aligned} \tag{11}$$

Since ρ_0 is necessarily unity and α_0 is taken as 1, we have

$$a_1 = -\frac{r_1(1 - r_2)}{1 - r_1^2} \quad \text{and} \quad a_2 = \frac{r_1^2 - r_2}{1 - r_1^2} \tag{12}$$

in which the a's are estimates of the α's and the r's are estimates of the ρ's This method, proposed originally by Yule [24], was examined empirically along with several alternatives and recommended by Kendall [8] for series of about 60 observations or more [3]

[3] In a paper by Rao and Som [17], one of the methods considered by Kendall [8] was found to be superior to (12) for series of the form (1) with $n=35$, $\alpha_1 = -0.7$ and $\alpha_2 = +0.6125$ This method consists of minimizing

$$\sum_{s=1} (r_{s+2} + a_1 r_{s+1} + a_2 r_s)^2$$

and involves the computation of r_s for several lags (in the example considered, r_1 through r_{11} were used). Since the Rao and Som paper came to the author's attention after the present study was essentially complete, the method is not examined extensively here. However, autoregression coefficients were estimated from the "average" segment (above) according to each of the two methods. Equation (12) leads to a's of $-.993$ and $+.577$ while the alternative method produces a's of $-.991$ and $+.561$. The results are very close and *both* methods are badly biased

Despite the Rao and Som results, there are strong reasons for preferring (12)

1 The shorter the series, the less favorable are the results we might expect from minimizing

$$\sum_{s=1} (r_{s+2} + a_1 r_{s+1} + a_2 r_s)^2,$$

values of r_s become more variable as s is increased, especially in the very short series in which we are interested

2. The methods as compared by Rao and Som were uncorrected for bias. Since both methods are biased, it is the efficacy of the correction for bias, not the superiority of the uncorrected method, that is the critical consideration.

3. The use of (12) requires considerably less computation.

The a's were computed from various r's using (12) and are shown in Table II. The deviations of the a's from their true values follow roughly the same pattern as those of the basic r's from their true values with one outstanding exception. The a's based on the average segment come much nearer to their respective parameters than do the r's from which they were computed There is no readily apparent reason for this.

The Effect of Correlation Between Segments. In the experimental series segments were independent of one another, but in actual series the 12 segments will usually be highly intercorrelated. Fortunately, this should not affect estimates based on the suggested methods.

For simplicity, we shall use Orcutt's [11] definition of r_s (with zero mean),

$$(13)\qquad \frac{n\sum_{i=1}^{n-s} x_i x_{i+s}}{(n-s)\sum_{i=1}^{n} x_i^2}.$$

This definition differs from that used above only with respect to "end-effects" and the generality of the exposition should not be affected. Let us compare the estimates r_s based on the series of annual totals (or annual averages) with those based on the whole series (segmented). Using (13),

$$(14)\qquad \frac{n\sum_{i=1}^{n-s} x_i x_{i+s}}{(n-s)\sum_{i=1}^{n} x_i^2} = \frac{n\sum_{j=1}^{12}\sum_{i=1}^{n-s} x_{ij}x_{i+s,j} + n\sum_{i=1}^{n-s}\sum_{j=1}^{12}\sum_{\substack{p=1-j\\ p\neq 0}}^{12-j} x_{ij}x_{i+s,j+p}}{(n-s)\sum_{j=1}^{12}\sum_{i=1}^{n} x_{ij}^2 + (n-s)\sum_{i=1}^{n}\sum_{j=1}^{12}\sum_{\substack{p=1-j\\ p\neq 0}}^{12-j} x_{ij}x_{i,j+p}},$$

$i=1, 2, \cdots, n$; $j=1, 2, \cdots, 12$; $p=1, 2, \cdots, (12-s)$; where

$$x_i = \sum_{j=1}^{12} x_{ij}.$$

Where the 12 series are independent, the terms to the right of the plus signs will tend toward zero and r_s based on the series of annual totals (or averages) will approximate that based on the whole series.

We shall show that, even where the 12 series are intercorrelated, ρ_s is estimated by the terms to the left of the plus signs in (14). We

TABLE II

SPECIFIED ESTIMATES OF AUTOCORRELATION AND AUTOREGRESSION COEFFICIENTS

Definition of Correlation Coefficient	Formula	n	Number of Series	Correlation Coefficient $s=1$	Correlation Coefficient $s=2$	α_1	α_2
ρ_s	*	∞	—	733	306	−1 1	.5
r_s, average segment	(7)	14	1	6301	0493	− 993	.577
$\bar{r}_{sj}$	(7), (6)	14	12	5533	0043	− .794	.435
$\bar{r}_{s(j,j+1)}$	(8), (5)	28	6	.6396	1516	− 918	436
$\frac{1}{2}({}_1r_s+{}_2r_s)$, method II	(10)†	84	2	6681	1960	− 970	452
r_s, whole series (seg)	(10)	168	1	6778	2310	− 964	.422
r_s (corrected, method I)	(4)	—	—	7259	2989	−1 076	482
r_s (corrected, method II)	(9)	—	—	6875	2660	− 957	392

* The ρ's that would be obtained from a series like (1) of infinite length with the α's shown are given by Kendall's [6, 8] formula,

$$\rho_s = \frac{|\sqrt{\alpha_2}|^s \sin(s\theta+\psi)}{\sin\psi},$$

where

$$\theta = \cos^{-1}\frac{-\alpha_1}{2\sqrt{\alpha_2}} \text{ and } \tan\psi = \frac{1+\alpha_2}{1-\alpha_2}\tan\theta$$

† In this case, formula (10) is applied separately to each of the two halves of the series rather than to the whole series as discussed in the text

have assumed that all segments have the same autocorrelation for each annual lag It is reasonable that,[9] asymptotically,

$$\text{(15)} \qquad \begin{cases} x_{i,j+p} = \beta x_{ij} + \epsilon_{i,j+p} \quad \text{and} \\ x_{i+s,j+p} = \beta x_{i+s,j} + \epsilon_{i+s,j+p} \end{cases}$$

where the β's are inter-segment linear regression coefficients and the ϵ's are uncorrelated with all x's and all other ϵ's. Substituting in (14) according to these relationships, we have

$$\text{(16)} \qquad \frac{n\sum_{i=1}^{n-s} x_i x_{i+s}}{(n-s)\sum_{i=1}^{n} x_i^2} = \frac{n\sum_{j=1}^{12}\sum_{i=1}^{n-s} x_{ij}x_{i+s,j} + n\sum_{i=1}^{n-s}\sum_{j=1}^{12}\beta_{jp}\sum_{\substack{p=1-j \\ p\neq 0}}^{12-j} x_{ij}x_{i+s,j}}{(n-s)\sum_{j=1}^{12}\sum_{i=1}^{n} x_{ij}^2 + (n-s)\sum_{i=1}^{n}\sum_{j=1}^{12}\beta_{jp}\sum_{\substack{p=1-j \\ p\neq 0}}^{12-j} x_{ij}^2}.$$

The β_{jp}'s are factors which leave the terms to the right of the plus signs proportionate to those to the left of the plus signs. Thus,

[9] This was suggested by Robert M Solow of Massachusetts Institute of Technology, one of the referees who reviewed this paper.

$$(17) \qquad \frac{n \sum_{i=1}^{n-s} x_i x_{i+s}}{(n-s) \sum_{i=1}^{n} x_i^2} \quad \text{and} \quad \frac{n \sum_{j=1}^{12} \sum_{i=1}^{n-s} x_{ij} x_{i+s,j}}{(n-s) \sum_{j=1}^{12} \sum_{i=1}^{n} x_{ij}^2}$$

have the same probability limit whether or not the segments are intercorrelated However, estimates made from short series according to the left side of (17) have the serious biases noted above Use of the much larger number of observations represented by the right side of (17) provides us with a vehicle for overcoming these biases, as currently proposed

Of course, the relationships (15) are asymptotic and it is not certain exactly how inter-segment correlation affects estimates based on very short series. A sampling study would be desirable

Testing Estimates Based on Actual Series While interdependence among the segments should not affect estimates of ρ_s, it does make it impossible to test the significance of coefficients computed according to the suggested methods from actual series Pending further study, the following approach may be considered a step toward independent segments and, eventually, methods of testing results in actual series. The approach requires a considerable amount of computation, but in view of the current rapid development of high-speed computers, this may not always be a serious difficulty.

If the original segments (months) are transformed by subtraction, the differences will be new segments that are often uncorrelated monthly but serially correlated annually [10] Since the existence of monthly serial correlation will destroy any tests of significance, it would seem necessary to compute the matrix of inter-month correlation coefficients so that the entire picture can be examined

Limitations of the Present Study It is clear that much remains to be done in the way of empirical investigation of the suggested approach

1 The definition of r_s developed by the author is only one of many probable ones. Other definitions may be better, and, as noted above, this possibility is examined by the author in a paper which is in progress.

2 The effect of seasonal variation on the present method has not been examined empirically. It is possible that much of the complexity due to seasonality could be cleared up by adjusting the original vari-

[10] Depending on the particular series, other expressions of monthly change (relative rather than absolute, etc.) might be appropriate.

ates for seasonal variation. However, it is not known to what extent the adjustments, themselves, affect the serial structure. The matter is one for further study, although the author suspects that the problem of seasonality is not too serious.

3. An efficient method of examining the efficacy of the present approach in short actual series has not yet been developed.[11]

4. The scheme (1) implies that all the error is incorporated into the system. In practice this is rarely, if ever, realized, the complications taking the form of "superposed variation"[12] or "errors in variables"[13] that may or may not be serially correlated. Attempts should be made to develop a useful method of estimation in short series where superposed variation or errors of observation exist. However, little is known about the real nature of superposed variation as it affects actual series so the amount to be learned via experimental series is problematical.[14] On the other hand, analyzing actual series presents the difficulty of unknown parameters, and a thorough study of superposed variation depends on the ability to derive the tests suggested in item 3 above.

5. The scheme (1) is not the only possible one nor necessarily the most useful. The choice of a particular model is important, since the bias considered above is a function not only of the sample size, as shown by Sastry [19], but also of the number of lags and the level of the α's Orcutt [11] or Cochrane and Orcutt [4] suggest a different scheme which might be examined along with other alternatives. A thorough examination would consider various α's and sample sizes for each of the model-types.

6. As noted in footnote 8, the method (12) may not be the best. The alternative method discussed has not been investigated in connection with the approach suggested in the present paper. Special attention should be given to the problem of choosing the optimum number of lags to use in the relationship to be minimized. Where two lags are used, the methods are identical.

7. The exact effect of inter-segment correlation on estimates based on very short series has not been examined. An empirical study is desirable.

---

[11] For large-sample tests of fit in autoregressive sequences, see Quenouille [15], Walker [22], Bartlett and Diananda [3], Rao and Som [17], and Rao [18] For considerations which may aid in the development of small-sample tests, see Sastry [20].

[12] See Kendall [6, 9] and Quenouille [15].

[13] See Stone [21] and Cochrane and Orcutt [4].

[14] For methods of detecting superposed variation or making estimates eliminating its effect, where a large number of lags can be considered, see Quenouille [15], Kendall [6] and Ghurye [5].

CONCLUSION

Of the limitations summarized above, the most important is that outlined in 4—superposed variation or errors of observation *not* incorporated into the autoregressive scheme. Where a series is long enough for the r's to be computed for many lags, methods of overcoming superposed variation are available.[12] However, for series of the length considered in the present study, these methods are not applicable and substitute methods are not known.

For the practical analyst to employ the approach suggested in this paper in his research, it will be necessary for him to consider first the extent to which his series are affected by superposed variation not incorporated into the generating process This examination will require great care since superposed random variation of even 5 or 10 per cent can be serious, and, at the current level of our knowledge, this examination must be mainly intuitive. Where "superposed variation" is serially correlated, it presents less difficulty (provided it extends over the life of the series), since it will then be incorporated into the autoregressive scheme. Fortunately, most errors of observation, etc., *are* serially correlated.[15] However, a superposed variation that is appreciable will obviously alter the basic scheme where its serial structure differs from that of the basic series

If the researcher realizes the limitations recorded above, he has available a crude tool where none was known before If the problem of superposed variation can be solved, the tool should become quite effective.

REFERENCES

[1] Anderson, T. W., "On the theory of testing serial correlation," *Skandinavisk Aktuarietidskrift*, 1948, 88–116.

[2] Bartlett, M S, "On the theoretical specification and sampling properties of autocorrelated time-series," *Supplement to the Journal of the Royal Statistical Society*, 8 (1946), 27–41

[3] Bartlett, M. S, and Diananda, P H, "Extensions of Quenouille's test for autoregressive schemes," *Journal of the Royal Statistical Society*, B-12 (1950), 108–15.

[4] Cochrane, D, and Orcutt, G. H., "Application of least squares regression to relationships containing autocorrelated error terms," *Journal of the American Statistical Association*, 44 (1949), 32–61.

[5] Ghurye, S G., "A method of estimating the parameters of an autoregressive time series," *Biometrika*, 37 (1950), 173–8.

[6] Kendall, Maurice G, *The Advanced Theory of Statistics*, II., London: Charles Griffin and Company, 1948, Chapter 30

[15] For examples of typical errors of observation and a brief discussion of the nature of their serial interdependence, see Cochrane and Orcutt [4].

[7] Kendall, M. G., "Contributions to the study of oscillatory time-series," *Occasional Paper* IX, National Institute of Economic and Social Research, Cambridge: Cambridge University Press, 1946.

[8] Kendall, Maurice G., "The estimation of parameters in linear autoregressive time-series," *Econometrica*, 17 (1949), July Meeting Supplement, 44–57.

[9] Kendall, M. G., "On autoregressive time-series," *Biometrika*, 33 (1944), 105–22.

[10] Moran, P. A. P., "Some theorems on time-series II The significance of the serial correlation coefficient," *Biometrika*, 35 (1948), 255–60.

[11] Orcutt, G H., "A study of the autoregressive nature of the time-series used for Tinbergen's model of the economic system of the United States, 1919–1932," *Journal of the Royal Statistical Society*, B-10 (1948), 1–53.

[12] Orcutt, Guy H., and Cochrane, Donald, "A sampling study of the merits of autoregressive and reduced form transformations in regression analysis," *Journal of the American Statistical Association*, 44 (1949), 356–72.

[13] Orcutt, G H., and James, S. F, "Testing the significance of correlation between time-series," *Biometrika*, 35 (1948), 397–413

[14] Quenouille, M. H., "Approximate tests of correlation in time-series," *Journal of the Royal Statistical Society*, B-11 (1949), 68–84.

[15] Quenouille, M H., "A large sample test for the goodness of fit of autoregressive schemes," *Journal of the Royal Statistical Society*, 110 (1947), 123–9.

[16] Quenouille, M H, "Some results in the testing of serial coefficients," *Biometrika*, 35 (1948), 261–7.

[17] Rao, S Raja, and Som, R. K., "The applicability of large sample tests for moving average and autoregressive schemes to series of short length—an experimental study, Part 2: autoregressive series," *Sankhyā: The Indian Journal of Statistics*, 11 (1951), 239–56

[18] Rao, C. Radhakrishna, "The applicability of large sample tests for moving average and autoregressive schemes to series of short length—an experimental study, Part 3. the discriminant function approach in the classification of time series," *Sankhyā The Indian Journal of Statistics*, 11 (1951), 257–72.

[19] Sastry, A Sree Rama, "Bias in estimation of serial correlation coefficients," *Sankhyā The Indian Journal of Statistics*, 11 (1951), 281–96.

[20] Sastry, A. Sree Rama, "Some moments of moment-statistics and their use in tests of significance in autocorrelated series," *Sankhyā The Indian Journal of Statistics*, 11 (1951), 297–308.

[21] Stone, Richard, "The role of measurement in economics," *Department of Applied Economics Monograph No. 3*, 1950. Cambridge Cambridge University Press

[22] Walker, A M., "Note on a generalization of the large sample goodness of fit test for linear autoregressive schemes," *Journal of the Royal Statistical Society*, B-12 (1950), 102–7.

[23] Wold, Herman, *A Study in the Analysis of Stationary Time-Series*, Uppsala. Almqvist & Wiksells Boktryckeri, 1938.

[24] Yule, G. Udny, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Philosophical Transactions*, Series A, 226 (1927), 267–98

ON GENERALIZATIONS OF TCHEBYCHEF'S INEQUALITY

H. J. Godwin
University College of Swansea, Wales

This paper is purely expository and contains an account of inequalities of the type of Tchebychef's, i.e., inequalities for the value of a distribution function in terms of known facts about the distribution. Such facts may be numerical, e.g., moments or range, or geometrical, e.g., the property of being unimodal or monotonic in some given range. The distributions to which they apply may be classified as single-variate, distributions of averages, or multivariate distributions. On pages 924–5 there is given a table of the results quoted, showing what data each uses and the type of distribution to which it applies. At the end of the paper are some remarks on possible future developments and on applications.

INTRODUCTION

If the frequency function $f(x)$ of a variate x is known, then we can calculate $P(a, b)$, the probability that $a \leqq x \leqq b$, exactly. If we know nothing about $f(x)$ then we can make only the trivial statement $0 \leqq P(a, b) \leqq 1$. Between these two extremes of complete knowledge and complete ignorance lie the situations in which we have some information about $f(x)$ and so can make a non-trivial statement about $P(a, b)$. The purpose of this paper is to give an account of the results of this kind which have been obtained

EXPECTATIONS AND MOMENTS

We shall assume throughout the paper that we are dealing with continuous frequency functions, except where discrete ones are specifically mentioned Consequently we write the expectation of the function $\Phi(x)$ as

$$E[\Phi(x)] = \int_{-\infty}^{\infty} \Phi(x) f(x) dx$$

whereas in the discrete case this would be

$$\sum_{i} \Phi(x_i) f(x_i)$$

We leave it to the reader to make the necessary modifications in formulas and statements to cover non-continuous cases The most serious

TABLE OF RESULTS QUOTED

The first column gives the number of the result below, the second the type of data and the third the type of region whose probability is being assigned bounds or the type of result if it is different from the general kind.

Distributions of Single Variates

(*i*)	Any *j* absolute moments $\nu_{i_1}, \cdots, \nu_{i_j}$.	Interval from origin.
(*ii*)	$\mu_1, \cdots, \mu_{2m}$	Semi-infinite interval.
(*iii*)	$\mu_1, \cdots, \mu_n$; finite range	Interval with one end at end of range.
(*iv*)	μ, μ_2.	Finite interval.
(*v*)	μ, μ_2, μ_4.	Two finite intervals, symmetrically placed about mean
(*vi*)	μ, μ_2, finite range.	Interval symmetrical about mean.
(*vii*)	Expectation of one function.	Finite interval
(*viii*)	Expectation of two functions.	Interval from origin.
(*ix*)	Mean range in samples of *n*.	*l.u.b.* of probability in all intervals of given length.
(*x*)	ν_r, graph of $P(0, x)$ lies below certain line in some range	Interval from origin.
(*xi*)	Conditions on derivatives in $a \leqq x \leqq b$, expectation of power of x over $a \leqq x \leqq b$, value of $f(x_0)$.	Interval $0 \leqq x \leqq x_0$
(*xii*)	As (*xi*) without $f(x_0)$.	Interval from origin.
(*xiii*)	As (*xii*)	Interval from origin
(*xiv*)	ν_{2r}; monotonicity condition on $f(x)$	Interval from origin
(*xv*)	ν_k, $f(x)$ has maximum at 0.	Interval from origin.
(*xvi*)	Discrete distribution; ν_k, monotonicity condition on probabilities.	Interval symmetrical about origin
(*xvii*)	$\mu_2(0)$, $f(x)$ has single maximum somewhere	Interval symmetrical about origin
(*xviii*)	ν_1, ν_2, mode at origin.	Interval from origin
(*xix*)	ν_{2r}, $f(x)$ has single maximum at given point	Interval from origin
(*xx*)	Mean range in sample of n, $f(x)$ is symmetrical about 0 and has single maximum	Interval symmetrical about 0.
(*xxi*)	ν_n, $f(x)$ non-decreasing in interval from 0	Interval from origin.
(*xxii*)	ν_n; $f(x)$ non-increasing in interval from 0.	Interval from origin.

Distributions of averages or totals

(*xxiii*)	μ, μ_2; sample estimate of μ_2.	Interval symmetrical about μ.
(*xxiv*)	μ, μ_2	Interval symmetrical about μ.
(*xxv*)	Restriction on all moments (or finite range)	Interval symmetrical about μ.
(*xxvi*)	As (*xxv*) but for cumulants.	Interval symmetrical about μ.
(*xxvii*)	μ_2, ν_3	*l.u.b.* of probability in interval of given length.
(*xxviii*)	μ, μ_2, ν_3	Semi-infinite interval.
(*xxix*)	μ, μ_2, ν_4	Semi-infinite interval
(*xxx*)	Symmetrical unimodal distribution; finite range.	Interval from origin
(*xxxi*)	Finite range, $\Sigma\mu_2$	Largest of cumulative totals in given interval

TABLE OF RESULTS QUOTED—*(continued)*

Multivariate distributions (first order moments zero)		
(xxxii)	Second order moments.	Certain rectangles.
(xxxiii)	Moments up to $2s^{\text{th}}$ order	Certain ellipses
(xxxiv)	Second order moments, independent variates.	Any ellipse with center at origin.
(xxxv)	Second order moments.	Certain quadrics or parallelepipeds.
(xxxvi)	Second order moments; monotonicity condition on distribution function	Ellipsoid with center at origin and axes parallel to coordinate axes.
(xxxvii)	Observed frequencies in independent trials	Bounds on probability of given size of measures of discrepancy of observed and true frequencies, or two lots of observed frequencies.
(xxxviii)	Contour moments.	Region inside contour

difference between the two kinds of distribution is that in the continuous case the probability that $a \leqq x \leqq b$ is the same as the probability that $a < x < b$, but in the discrete case, if there is positive probability of $x = a$ or $x = b$, the two expressions are different.

The most extensive results are those which depend on a knowledge of some moments of the distribution Since the notation used for moments varies somewhat we state explicitly the one to be used here.

If a is any value we put $E\left[(x-a)^n\right] = \mu_n(a)$, and for absolute moments $E(|x-a|^n) = \nu_n(a)$. We then have $\mu_0(a) = \nu_0(a) = 1$ for all a. We write μ for the mean value $\mu_1(0)$ and write $\mu_n(\mu)$ simply as μ_n. In the case of absolute moments we shall always measure the variate from the point about which moments are taken, which means that we take a to be 0, and then denote the moment by ν_n

It should be noted that ν_n is the same for the frequency functions $f(x)$ and $g(x)$, where $g(x) = 0$ for $x < 0$, and $g(x) = f(x) + f(-x)$ for $x \geqq 0$. The probability which is $P(-d, d)$ for $f(x)$ is $P(0, d)$ for $g(x)$ and it will be convenient to state results for absolute moments always in terms of distributions of non-negative variates. As long as we are concerned with probabilities $P(-d, d)$ over ranges symmetrical about the origin, there will be no loss of generality.

RESULTS DEPENDING ON MOMENTS

Suppose that we are given for $f(x)$ either $(i) j$ absolute moments $\nu_{i_1}, \cdots, \nu_{i_j}$ (where $i_1 < i_2 < \cdots < i_j$) or (ii) the $2n$ moments $\mu_1, \cdots, \mu_{2n}$ or (iii) the n moments $\mu_1, \cdots, \mu_n$ when $f(x)$ is zero outside the interval $b \leqq x \leqq c$. Then bounds for $P(0, d)$, $P(-\infty, d)$, $P(b, d)$ in these cases respectively can be found as follows. We construct discrete distributions of certain kinds whose moments equal the corresponding

ones of $f(x)$; each of these distributions has non-zero probabilities at only a finite number of points, which constitute the spectrum of the distribution. In the three cases the spectra consist of the following points.

(i) if j is even; d and not more than $\frac{1}{2}j$ other positive points, *or* 0, d, not more than $\frac{1}{2}(j-2)$ other positive points, and ∞. (At ∞ is placed zero probability in such a way that ν_j is affected but no moment of lower order.)

(i) if j is odd; 0, d and not more than $\frac{1}{2}(j-1)$ other positive points.

(ii) d and n other points.

(iii) if n is even; d and not more than $\frac{1}{2}n$ other points in the interval $b<x<c$, *or* b, c, d, and not more than $\frac{1}{2}(n-2)$ other points in the interval $b<x<c$.

(iii) if n is odd; b, d and not more than $\frac{1}{2}(n-1)$ other points in the interval $b<x<c$, *or* d, c and not more than $\frac{1}{2}(n-1)$ other points in the interval $b<x<c$

It has been shown that in each case just one construction is possible, except that for a finite number of values of d, depending on the given moments, we have to replace d by $d+\delta$ and let δ tend to 0 to obtain the spectrum and associated probabilities

We have now constructed in each case discrete distributions, consisting of a finite number of points and d, with certain probabilities at each. The lower bounds for the probabilities $P(0, d)$, $P(-\infty, d)$ and $P(b, d)$ are the sums of the probabilities in the discrete distributions at points to the left of d, and the upper bounds the sum of probabilities at points to the left of d and at d. Since these bounds are given arbitrarily closely by distributions departing by infinitesimal amounts from the discrete distributions which have been constructed, the inequalities cannot be improved on except by excluding distributions of this discrete type, we shall return to this point later. Inequalities which cannot be improved on, under the given conditions, are called best possible

A proof of the above results in case (i) is given by Wald [38] and outlines of the proofs in cases (ii) and (iii), with further references, by Shohat and Tamarkin [35, pp. 43 and 79].

To show the method in operation let us take first the case where we know the one absolute moment ν_n We have to find a discrete distribution having probabilities p at 0 and q at d and with this moment. Hence we must have

$$p + q = 1, \qquad qd^n = \nu_n,$$

whence

$$p = 1 - \nu_n/d^n, \qquad q = \nu_n/d^n$$

and

$$1 - \nu_n/d^n \leqq P(0, d) \leqq 1.$$

The right-hand inequality is trivial, but for $d^n > \nu_n$ the left-hand one is not.

If $n=1$ we have $1-\nu_1/d \leqq P(0, d)$, which is known as Markoff's Lemma.

If $n=2$, since $\mu_2(0)=\nu_2$ we can write the inequality as $1-\mu_2(0)/d^2 \leqq P(-d, d)$. If the standard deviation of the distribution is σ then, by a change of origin, we can write the inequality as $1-t^{-2} \leqq P(\mu - t\sigma, \mu + t\sigma)$. This inequality, discovered by Bienaymé in 1853 and rediscovered by Tchebychef in 1867 is often called Tchebychef's inequality and gives the name to the whole subject

As a second example let ν_n, ν_{2n} be known. Then we construct a distribution with either (A) probability p at d and $1-p$ at e, or (B) probabilities p at d, $1-p$ at 0 and 0 at ∞. We require (A) $pd^n+(1-p)e^n = \nu_n$, $pd^{2n}+(1-p)e^{2n}=\nu_{2n}$, or (B) $pd^n=\nu_n$, $pd^{2n} \leqq \nu_{2n}$. Eliminating e^n in (A) we have $(\nu_n - pd^n)^2 = (1-p)(\nu_{2n}-pd^{2n})$, whence $p=(\nu_{2n} - \nu_n^2)/(\nu_{2n}-\nu_n^2+(d^n-\nu_n)^2)$, $e^n=(\nu_n d^n - \nu_{2n})/(d^n-\nu_n)$ In (B) we have $p=\nu_n/d^n$ Since, by Schwartz's inequality, $\nu_{2n} \geqq \nu_n^2$ we have $0 \leqq p \leqq 1$ in (A), but the condition $e^n \geqq 0$ requires either $d^n \leqq \nu_n$ or $d^n \nu_n \geqq \nu_{2n}$ In (B) we have $\nu_{2n} \geqq d^n \nu_n$ and we need $\nu_n \leqq d^n$ so that $p \leqq 1$ Thus for a given set of values of d, ν_n, ν_{2n} just one of (A) or (B) applies. (Except that if $\nu_n = d^n$ or $\nu_n d^n = \nu_{2n}$ we have limiting cases which belong to either (A) or (B)).

If $d^n \leqq \nu_n$ then in (A) we have $e \geqq d$ so that

$$0 \leqq P(0, d) \leqq p = \frac{\nu_{2n} - \nu_n^2}{\nu_{2n} - \nu_n^2 + (d^n - \nu_n)^2}.$$

If $\nu_n \leqq d^n \leqq \nu_{2n}/\nu_n$ then, from (B),

$$1 - p = 1 - \nu_n/d^n \leqq P(0, d) \leqq 1.$$

If $d^n \geqq \nu_{2n}/\nu_n$ so that $d^n \geqq \nu_n$ then, in (A), we have $e \leqq d$ so that

$$1 - p = \frac{(d^n - \nu_n)^2}{\nu_{2n} - \nu_n^2 + (d^n - \nu_n)^2} \leqq P(0, d) \leqq 1.$$

These inequalities were first given by Cantelli in 1928.

For some solutions in the case (ii) see Zelen [40].

A simple extension of the inequality using one moment was made in its most general form by Lurquin [20]; if the events $e_1, \cdots, e_k$ occur with probabilities $p_1, \cdots, p_k$ and if with event e_i is associated the measurement x_i with absolute moment $\nu_n(i)$ then the probability that for at least one i we have $|x| < d$ is greater than $1-(\sum_i p_i \nu_n(i)/d^n$.

Since the knowledge of a number of moments gives bounds for $P(-\infty, d)$ it follows that if a number of corresponding moments of two frequency functions $f_1(x)$ and $f_2(x)$ are equal, then the respective probabilities $P_1(-\infty, d)$ and $P_2(-\infty, d)$ will not differ widely. Bounds for $|P_1(-\infty, d)-P_2(-\infty, d)|$ in terms of the moments have been given by a number of authors. Khamis [16] quotes some of these, a simple form being

$$|P_1(-\infty, d) - P_2(-\infty, d)| \leq \begin{vmatrix} \mu_0(a) & \cdots & \mu_n(a) \\ \cdot & \cdots & \cdot \\ \mu_n(a) & \cdots & \mu_{2n}(a) \end{vmatrix} \Bigg/ \begin{vmatrix} \mu_2(d) & \cdots & \mu_{n+1}(d) \\ \cdot & \cdots & \cdot \\ \mu_{n+1}(d) & \cdots & \mu_{2n}(d) \end{vmatrix},$$

where a is any value. Khamis also improves this result in terms of a knowledge of bounds for the ratio $f_1(x)/f_2(x)$.

We have so far considered $P(a, b)$ only for the cases in which a is at the extreme left-hand end of the distribution. Inequalities for general intervals can be derived by subtraction from the inequalities which we have given, but will now no longer, in general, be best possible. Selberg [33] has however given a best possible result in terms of μ_2 for any interval including the mean value. The result is

(iv)

$$P(\mu - \alpha, \mu + \beta) \geqq \alpha^2/(\alpha^2 + \mu_2) \quad \text{for} \quad \alpha(\beta - \alpha) \geqq 2\mu_2;$$

$$P(\mu - \alpha, \mu + \beta) \geqq (4\alpha\beta - 4\mu_2)/(\alpha + \beta)^2 \text{ for } 2\alpha\beta \geqq 2\mu_2 \geqq \alpha(\beta - \alpha),$$

where $\beta \geqq \alpha \geqq 0$.

For $\mu_2 \geqq \alpha\beta$ we have only the trivial inequality

$$P(\mu - \alpha, \mu + \beta) \geqq 0.$$

Guttman [13] shows that

(v)

$$P(\mu - k_1\sigma, \mu - k_2\sigma) + P(\mu + k_2\sigma, \mu + k_1\sigma) \geqq 1 - \lambda^{-2},$$

where $k_1 = (1+\lambda\alpha)^{1/2}$, $k_2 = (1-\lambda\alpha)^{1/2}$ or $k_1 = (1+\alpha(\lambda^2+1)/\sqrt{(\lambda^2-1)})^{1/2}$, $k_2 = (1-\alpha\sqrt{(\lambda^2-1)})^{1/2}$, or $k_1 = (1+\alpha\sqrt{(\lambda^2-1)})^{1/2}$, $k_2 = (1-\alpha(\lambda^2+1)/$

$\sqrt{(\lambda^2-1)})^{1/2}$ and $\sigma^2=\mu_2$, $\mu_4=(\alpha^2+1)\mu_2$. If the expression for k_2^2 is negative we replace k_2 by 0; for the result to be best possible we require respectively $\alpha\lambda\leqq 1$, $\alpha\sqrt{(\lambda^2-1)}\leqq 1$ or $\alpha\leqq\sqrt{(\lambda^2-1)}$, the distribution for which equality is obtained being a discrete one whose spectrum has not more than six points.

For distributions of finite range the above inequalities can be improved on. If z is the maximum value of $|x-\mu|$ then Lurquin [21] gives

(vi) $$P(\mu - t\sqrt{\mu_2}, \mu + t\sqrt{\mu_2}) \leqq 1 - \mu_2(1 - t^2)/z^2.$$

This is not best possible but can immediately be improved to $P(\mu-t\sqrt{\mu_2}, \ \mu+t\sqrt{\mu_2})\leqq 1-\mu_2(1-t^2)/(z^2-t^2\mu_2)(t<1)$ which is best possible.

RESULTS DEPENDING ON EXPECTATIONS OF GENERAL FUNCTIONS

As a generalization of knowledge of moments of a distribution we may suppose a knowledge of the expectations of more general functions; results are available only when the number of functions is one or two.

(vii) Let $g(x)$ be a positive function of x, increasing for $x\geqq d\geqq 0$. Then if $m_g=\int_0^\infty\{f(x)+f(-x)\}g(x)dx$ we have

$$P(-d, d) \geqq 1 - m_g/g(d).$$

This seems to have been first stated by Cantelli [9].

The following result for the case when the expectations of two functions are known was established in a very elegant way by von Mises [24]

(viii) Let $g(x)$, $h(x)$ be two increasing functions of x for positive x, such that $g(0)=h(0)=0$ and the curve Γ given by the parametric equations $x=g(t)$, $y=h(t)$ is concave upwards (i.e., having a tangent whose gradient increases with x). Let the frequency function of x be zero for negative x and let $g=E(g(x))$, $h=E(h(x))$. In the (x, y) plane let C be the point (g, h), $Q(p, q)$ the point of Γ for which t is largest (this is possibly a point at infinity for which only the ratio p/q is finite), t_1 the parameter of the point at which OC meets Γ, t_2 the parameter of the point at which QC meets Γ, M the point of Γ with parameter d, and let CM meet Γ in the point (g', h'). Then

$$\text{for } d \leqq t_2, \qquad 0 \leqq P(0, d) \leqq 1 - \frac{g - g(d)}{g' - g(d)};$$

$$\text{for } t_2 \leqq d \leqq t_1, \qquad \frac{(g - p)(h - h(d)) - (g - g(d))(h - q)}{ph(d) - qg(d)}$$

$$\leqq P(0, d) \leqq 1 - \frac{gh(d) - hg(d)}{ph(d) - qg(d)};$$

for $t_1 \leqq d$, $$\frac{g - g(d)}{g' - g(d)} \leqq P(0, d) \leqq 1.$$

The above inequalities are best possible, being satisfied with equality for certain distributions with not more than three different values of x.

A RESULT USING MEAN RANGE

The mean range in samples of n from the population is a function which is not covered by any of the above results Winsten [39] has used it to obtain an inequality as follows

(ix) Let w_n be the mean range in a sample of n; let t be given and m determined by

$$\sum_{i=1}^{m}\left\{1 - \left(\frac{i}{m}\right)^n - \left(1 - \frac{i}{m}\right)^n\right\} \leqq \frac{1}{t} < \sum_{i=1}^{m+1}\left\{1 - \left(\frac{i}{m+1}\right)^n - \left(1 - \frac{i}{m+1}\right)^n\right\}.$$

Let p be determined by

$$t^{-1} = \sum_{i=1}^{m} \{1 - (ip)^n - (1 - ip)^n\}$$

Then

$$\operatorname*{l.u.b.}_{x} P(x, x + tw_n) \geqq p.$$

This inequality is different in character from the preceding ones, since it relates not to a single value of $P(a, b)$ but to the least upper bound of a certain set of values

The inequality is replaced by equality for a certain distribution with $(m+1)$ distinct values

DISTRIBUTIONS WITH GEOMETRICAL RESTRICTIONS

All the inequalities which we have so far considered can be replaced by equality for distributions with probability at a finite number of points. We can improve the inequalities by excluding such distributions and this we do by restricting the number of maxima of the frequency function, or by imposing conditions of monotonicity or convexity over various parts of the range of the variate. Most of the results

which follow are also best possible, equality being attained when the graph of the frequency function consists of rectangular blocks, possibly with extra positive probabilities at certain points superimposed.

A fairly general condition is that of von Mises [23]. We replace a general distribution by one with non-zero x as explained in the introductory remarks about absolute moments and require that for $x>x_0$ the graph of $P(0, x)$ shall lie below the line joining $(z, 1)$ and $(x_1, P(0, x_1))$. For a given absolute moment ν_r we define ζ, t by

$$(r + 1)\zeta^r(\zeta - x_1) = \zeta^{r+1} - x_0^{r+1}$$

$$(r + 1)\nu_r(\tau - x_0) = \tau^{r+1} - x_0^{r+1}.$$

(x) Then

$$P(0, x_1) \geqq 1 - \nu_r/\zeta^r$$

and if $\nu_r > x_0^r$ and $\zeta < \tau$ then we have also

$$P(0, x_1) \geqq (x_1 - x_0)/(\tau - x_0).$$

A particular case of von Mises' condition is when $f(x)$ is monotonically decreasing in some range; more generally than this we may require some derivative of $f(x)$ to vary monotonically. Van Dantzig [12] proved the following result

(xi) Let x be a non-negative variate and denote $P(x, \infty)$ by $R(x)$ In $a \leqq x \leqq b$, where $0 \leqq a < b < \infty$, let $(-1)^j R^{(j)}(x)$ be positive for $0 \leqq j \leqq h$ and non-increasing for $j = h$. Let h^* be the integral part of $(b - x_0)f(x_0)/P(x_0, b)$, $H = \min(h, h^*)$, $r = x_0 + HP(x_0, b)/f(x_0)$,

$$\alpha_k' = \int_a^b (x^k - a^k)f(x)dx, \qquad \Gamma_{H,k}(\alpha) = \max_{0 \leq \rho \leq 1} \frac{\rho^k(1 - \rho)^H}{k\int_{\alpha\rho}^1 \xi^{k-1}(1 - \xi)^H d\xi}.$$

Then

$$P(0, x_0) \geqq P(0, b) - \Gamma_{H,k}(a/r)\alpha_k'/x_0^k.$$

Except when $a = 0$ this involves $P(0, x_0)$ on both sides of the inequality, as well as using $f(x_0)$ A less powerful result, which avoids these disadvantages, is

(xii)

$$P(0, x_0) \geqq P(0, b) - \left\{ \Phi_{H,k}\left(\frac{a}{x_0}, \frac{\int_a^b x^k f(x)dx}{x_0^k P(a, b)} \right) \right\}^H P(a, b),$$

where $\Phi(\alpha,C)$ satisfies

$$\Phi^H/\gamma_{H,k} - \sum_{j=1}^{H}(-1)^{j-1}\frac{H!k^j}{(H-j)!(k+j)!}\Phi^{H-j}\left(\frac{1-\phi}{1-\alpha}\right)^j I_{k+1,j}(\alpha)=C,$$

$$I_{l,H}(\alpha) = \sum_{j=0}^{H-1} {}^{l+H-1}C_{l+j}\alpha^{l+j}(1-\alpha)^{H-1-j}$$

and

$$\gamma_{H,k} = \Gamma_{H,k}(0) = \frac{(H+k)!k^kH^k}{H!k!(H+k)^{H+k}}\cdot$$

More simply still we have
(xiii)

$$P(0, x_0) \geqq P(0, b) - \gamma_{H,k}\frac{\int_a^b x^kf(x)dx}{x_0{}^k(1+\lambda)} - \frac{P(a, b)\psi}{1+\lambda^{-1}},$$

where

$$\lambda = \frac{k^kH^H}{(k+H)^{k+H}}\;\frac{(a/x_0)^{k+H}}{(1-a/x_0)^H}, \qquad \psi = \frac{I_{k+1,H}(a/x_0)}{(a/x_0)^{H+k}}\cdot$$

When $h=1$ the results in (xii) and (xiii) are the same.

If in (xiii) we put $h=1$, $k=2r$, $b=\infty$ and neglect $\int_0^a x^kf(x)dx$, we obtain

(xiv) $$P(0, x_0) \geqq 1 - \theta - \nu_{2r}/((2r+1)/2r)^{2r}(1+\lambda)x_0{}^{2r}$$

where

$$\theta = P(0, a)/(1+\lambda^{-1}), \qquad \lambda = \left(\frac{a}{x_0}\,\frac{2r}{2r+1}\right)^{2r}/(2r+1)\left(\frac{x_0}{a}-1\right).$$

This is an earlier result proved by Camp in 1922. If $a=0$ and $h=1$ we have

(xv) $$P(0, x_0) \geqq 1 - \frac{\nu_k}{x_0{}^k}\left(\frac{k}{k+1}\right)^k,$$

which was proved by Meidell in 1922. Meidell also [22] showed that (xvi) if x takes the values A, $A+1, \cdots, B$ with probabilities $f(A), \cdots, f(B)$ and if, in the interval $|x| \leqq c(1+n^{-1})$ these probabili-

ties have a single maximum at 0, then the sum of the probabilities in the interval $|x| \leqq c$ is not less than

$$1 - \frac{\nu_n}{c^n}\left(\frac{n}{n+1}\right)^n.$$

A particular case of (xv) is $P(0, d) \geqq 1 - 4\nu_2/9\alpha^2$, or, for a distribution with a variate not necessarily positive,

$$P(-d, d) \geqq 1 - 4\mu_2(0)/9d^2;$$

the condition on the frequency function is that it should have a maximum at $x=0$. This result was conjectured by Gauss.

Selberg [32] has shown that

(xvii) if $f(x)$ has a single maximum not necessarily at $x=0$, then

$$P(-d, d) \geqq 1 - \theta\mu_2(0)/d^2,$$

where

$$= .565376 \cdots \quad \text{satisfies } \theta^3 - 9\theta^2 + 3\theta + 1 = 0.$$

(This result is not best possible). He also [34] gives a more complicated result in terms of other moments

Royden [30] gives inequalities for a unimodal distribution in terms of the first and second absolute moments about the mode (which we take to be at the origin) The results are:

(xviii) $$P(0, d) \leqq 1 - \frac{(2\nu_1 - d)^2}{3\nu_2 - 2d\nu_1} \quad \text{for} \quad 0 \leqq d \leqq 2\nu_1,$$

$$P(0, d) \leqq 1 \quad \text{for} \quad 2\nu_1 \leqq d,$$

and

$$P(0, d) \geqq d/2\nu_1 \quad \text{for} \quad 0 \leqq d \leqq \nu_1,$$

$$P(0, d) \geqq 1 - \nu_1/2d \quad \text{for} \quad \nu_1 \leqq d \leqq 3\nu_2/4\nu_1,$$

$$P(0, d) \geqq 1 - \frac{4\nu_1^2}{3\nu_2} + \frac{8\nu_1^3 d}{9\nu_2^2} \quad \text{for} \quad \frac{3\nu_2}{4\nu_1} \leqq d \leqq \frac{\nu_2}{\nu_1},$$

$$P(0, d) \geqq 1 - \frac{\Lambda - 1}{3t^2 - 4t + \Lambda} \quad \text{for} \quad \frac{\nu_2}{\nu_1} \leqq d,$$

where $\Lambda = 3\nu_2/4\nu_1^2$ (and is necessarily $\geqq 1$) and

$$d = 4\nu_1 \frac{t^3 - t^2}{3t^2 - 4t + \Lambda}.$$

(By a slip, Royden omits a 2 in the first inequality and has $3\nu_2 - d\nu_1$).

Smith [36, 37] considers the distribution of a non-negative variate whose frequency function has a single maximum at c and obtains

$$\text{(xix)} \qquad P(0, d) \geqq 1 - \frac{\nu_{2r} - c^{2r}(1 - 2rc/d(2r + 1))}{(d/\theta)^{2r} - c^{2r}(1 - 2rc/d(2r + 1))}$$

$$\text{for} \quad d < c/P(0, c),$$

where θ is defined by $d = 2r\{d^{2r+1} - (c\theta)^{2r+1}\}/(2r+1)\theta\{d^{2r} - (c\theta)^{2r}\}$.

If the frequency function of x has a single maximum at $x=0$ and is symmetrical about $x=0$ then Winsten [39] shows that

$$\text{(xx)} \qquad \begin{aligned} &P(-tw_n, tw_n) \\ &\geqq 1 - 4t\left\{\frac{1 + y^{n+1} - (1 - y)^{n+1}}{n + 1} - y^{n+1} - y(1 - y)^n\right\}, \end{aligned}$$

where $\dfrac{1}{2t} = 1 - (1-y)^n - y^n$ and w_n is the mean range in a sample of n.

Narumi [25] considers the case of the frequency function of a non-negative variate which is monotonic in a range starting at the origin (previous results of this kind have imposed conditions on the tail of the distribution)

(xxi) If $f(x)$ is non-decreasing in $0 \leqq x \leqq b$, where $\nu_n < b^n < (n+1)\nu_n$, then

$$0 \leqq P(0, d) \leqq d/b \qquad \text{for} \quad 0 \leqq d \leqq b_1,$$

$$\frac{d}{b} - \frac{b - d}{b^n - d^n}\,\frac{(n + 1)\nu_n - b^n}{b} \leqq P(0, d) \leqq \frac{d}{b} \quad \text{for} \quad b_1 \leqq d \leqq b,$$

$$1 - \frac{(n + 1)\nu_n - b^n}{(n + 1)d^n - b^n} \leqq P(0, d) \leqq 1 \quad \text{for} \quad b \leqq d,$$

where b_1 is the positive root of $d(b^n - d^n)/(b-d) = (n+1)\nu_n - b^n$ For $b^n = (n+1)\nu_n$ these results have the simple limiting form $P(0, d) = d/b$ for $0 \leqq d \leqq b$, and $P(0, d) = 1$ for $b \leqq d$. The case $b^n > (n+1)\nu_n$ is impossible.

(xxii) If $f(x)$ is non-increasing in $0 \leqq x \leqq b$, where $b^n > \nu_n$ then

$$\frac{(n + 1)d}{nb}\left(1 - \frac{\nu_n}{b^n}\right) \leqq P(0, d) \leqq 1 \quad \text{for} \quad 0 \leqq d \leqq nb/(n + 1),$$

$$1 - \frac{\nu_n}{b^n} \leqq P(0, d) \leqq 1 \quad \text{for} \quad nb/(n+1) \leqq d \leqq b,$$

$$1 - \frac{\nu_n}{d^n} \leqq P(0, d) \leqq 1 \quad \text{for} \quad b \leqq d.$$

DISTRIBUTIONS OF AVERAGES OR TOTALS

We have so far used information about the distribution of x, but nothing about the nature of x. We shall now consider a group of inequalities which deal with the case in which x is a sum of values, each with a probability distribution Many of the above inequalities can now be improved upon since for sufficiently large values of n the sum of n values cannot take a small number of discrete values. Some of the impetus for work in this field has been the Law of Large Numbers, which deals with the limiting form of the distribution of a sum, as the number of components becomes large, and several of the inequalities obtained, though adequate for their application to this theory, involve unknown constants which render them useless from the practical point of view. None of the inequalities which follow is best possible, except possibly for some special values of the parameters.

We write $Q(a, b)$ for the probability of the inequality $a \leqq \bar{x} \leqq b$, where $\bar{x}$ is the average of the n values added, and any moments or other data used in the definition of a or b are those of the parent populations We assume that the values which are added are independent unless otherwise stated.

Guttman [14] showed that

(xxiii)
$$Q\left(\mu - \sqrt{\left\{\frac{s^2}{n-1} + \mu_2\sqrt{\left(\frac{2(\lambda^2-1)}{n(n-1)}\right)}\right\}}, \right.$$
$$\left. \mu + \sqrt{\left\{\frac{s^2}{n-1} + \mu_2\sqrt{\left(\frac{2(\lambda^2-1)}{n(n-1)}\right)}\right\}}\right) \geqq 1 - \frac{1}{\lambda^2},$$

where

$$s^2 = \frac{1}{n}\sum (x_i - \bar{x})^2$$

and the x's are from the same population

Robbins [28] showed that

(xxiv) if $\mu=0$, $\mu_2=1/n$ and l b. $Q(-t/n, t/n)$ over all distributions (the x's coming from the same distribution) is $1-\phi_n(t)$ then $\phi_n(t)$

$<t^{-2}$ (strict inequality) for $n>1$ and $t>\sqrt{n}$ but that $t^2\phi_n(t)\to 1$ as $t\to\infty$ for all n. He proposed the interesting problem of finding

$$\overline{\lim_{n\to\infty}}\, \phi_n(t)$$

for all t.

Bernstein [3] imposed the following condition on the moments of each of the distributions from which values $x_1, \cdots, x_n$ are taken, $|\mu_r| \leqq \frac{1}{2}H^{r-2}r!\, \mu_2$ for all $r\geqq 2$, and some constant H. Then if $\mu=0$ for each of $x_1, \cdots, x_n$ and B_n is the variance of $x_1+\cdots+x_n$ so that $B_n=\sum\mu_2$, we have

$$\text{(xxv)} \qquad Q(-\lambda/n, \lambda/n) \geqq 1 - 2exp\frac{-\lambda^2}{2B_n+2H\lambda}.$$

If the x's have finite range so that $|x_i| \leqq M$ then we may take $H=M/3$.

By using the idea of conditional expectation (e.g., expectation of x_k^r relative to given values of $x_1, \cdots, x_{k-1}$) he was able in [4] to remove the condition that $x_1, \cdots x_n$ be independent.

Craig [11] replaced ordinary moments by moments over some finite interval $(-b, b)$, where b is arbitrarily large, and showed that a further possible condition is

$$\text{(xxvi)} \qquad |\lambda_r| \leqq \tfrac{1}{2}H^{r-2}r!\lambda_2,$$

where λ_r is the rth cumulant (again over a finite interval if desired)

Offord [26] obtained a lower bound for probability in an interval as follows, let the moments of $x_1, \cdots, x_n$ satisfy

$$\text{(xxvii)} \qquad \mu_2^{1/2}/\nu_3^{1/3} \geqq 2k^{1/3}.$$

Then

$$\operatorname*{l.u.b.}_{t} Q\left(\frac{t-x}{n}, \frac{t+x}{n}\right) \leqq \frac{6 \log n}{k^3\sqrt{n}}\left(\log n + \frac{kx}{\min \mu_2^{1/2}}\right),$$

the min in the bracket being over the distributions of $x_1, \cdots, x_n$. He also showed that the n, k, min $\mu_2^{1/2}$ can be replaced by the same quantities derived from a subsequence of at least two terms from $x_1, \cdots, x_n$.

As n increases, the distribution of $\bar{x}$, under certain conditions, tends to normality and so it is possible to estimate $Q(a, b)$ in terms of the

normal probability integral. If for the distribution of x_i we have $\mu=0$, $\mu_2=\mu_2(i)$, $\nu_3=\nu_3(i)$ and

$$B_n = \sum_{i=1}^{n} \mu_2(i),$$

then Berry [5] proved that

$$\text{(xxviii)} \quad \left| Q\left(-\infty, \frac{\lambda\sqrt{B_n}}{n}\right) - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\lambda} e^{-t^2/2}dt \right| < \frac{1.88}{\sqrt{B_n}} \max_i \frac{\nu_3(i)}{\mu_2(i)}.$$

Berry's paper is not without errors, as was remarked by Hsu [15, p. 3] but a referee tells me that the constant 1 88 is claimed to be still correct after the errors have been rectified. Bergström [2] proved a similar result with right-hand side

$$\text{(xxix)} \qquad (4.8) \sum_{i=1}^{n} \nu_3(i)/(B_n)^{3/2},$$

which is less good than Berry's for the case when the x's have identical distributions but is stronger in certain cases when they have not. Results of this type can also be proved using absolute moments other than the third order one, and for multivariate distributions, but then the constants which occur have not been evaluated numerically.

By comparing a variate with a finite range with a rectangularly distributed variate, Birnbaum [7] obtained

(xxx) if the x's have the same symmetrical, unimodal distribution with range $2a$ then

$$Q(-d, d) \geqq 1 - \Psi_n(d/a),$$

where

$$\Psi_n(t) = \frac{2}{n} \sum_{\frac{1}{2}n(t+1)<k\leqq n} (-1)^k\, {}^nC_k \left\{\frac{n}{2}(t+1) - k\right\}^n.$$

Some inequalities involving cumulative sums are due to Kolmogoroff [17, 18]. Here we consider not only the average S_n/n of n quantities $x_1, \cdots, x_n$ but also the sub-totals $S_r(=x_1+\cdots+x_r)$ and find bounds for the probability that at least one of these exceeds a given amount, or in other words that the largest one does.

(xxxi) If M is $\max_{1\leq i\leq n} |x_i-E(x_i)|$, u is $\max_{1\leq r\leq n} |S_r-E(S_r)|$ and

$D^2=E(S_n-E(S_n))^2$ then $Pr(u\geqq R)\leqq D^2/R^2$ while $Pr(u\geqq mR)\leqq D/(R-M)^{2m}$ if m is an integer and $R>M$. Also $Pr(\max_{1\leqq r\leqq n}|S_r|\leqq R)\leqq 4(M+R)^2/D^2$. For $D\geqq M$ we have $Pr(|S_n|\geqq\frac{1}{2}D)\geqq 1/1600$, (i.e., $Q(-\infty,-D/2n)+Q(D/2n,\infty)\geqq 1/1600$ in the notation used earlier) and from this follows trivially

$$Q(-\infty,\ -R/n)+Q(R/n,\ \infty)\geqq\frac{1}{1600}\left(1-\frac{M^2+4R^2}{D^2}\right)$$

for any R, M, D. The constant 1/1600 could be improved on even without departing from Kolomgoroff's proof.

Other results on means involve unknown constants and will not be given here.

MULTIVARIATE DISTRIBUTIONS

Some work has been done in the direction of extending the above results to n-variate distributions for $n>1$; the problem is now to find bounds for the amount of probability falling in some region of space of n dimensions Owing to the different kinds of region which are possible we can use no unifying notation as in the case of one dimension where only intervals have to be considered.

We suppose that the mean value of each variate is zero and write $\mu_{r\cdots t}=E(x_1^r\cdots x_n^t)$ As a special case for $n=2$ it is convenient to write $\mu_{20}=\sigma_1^2$, $\mu_{02}=\sigma_2^2$, $\mu_{11}=\rho\sigma_1\sigma_2$.

(xxxii) Berge [1] has shown that the probability that (x_1, x_2) falls in the rectangle $|x_1|<k\sigma_1$, $|x_2|<k\sigma_2$ is not less than $1-(1+\sqrt{(1-\rho^2)})/k^2$ and that this is best possible

(xxxiii) Pearson [27] showed that the probability that (x_1, x_2) falls in the ellipse $\theta_{11}x_1^2/\sigma_1^2-2\theta_{12}x_1x_2/\sigma_1\sigma_2+\theta_{22}x_2^2/\sigma_2^2=\lambda_0^2(1-\rho^2)$ is not less than $1-I_s/\lambda_0^{2s}$ where

$$I_s=\iint\frac{1}{1-\rho^2}\left(\theta_{11}\frac{x_1^2}{\sigma_1^2}-2\theta_{12}\frac{x_1x_2}{\sigma_1\sigma_2}+\theta_{22}\frac{x_2^2}{\sigma_2^2}\right)f(x_1,\ x_2)dx_1dx_2,$$

this integral being expressible in terms of moments of the distribution. The result is best possible, equality being obtained when the probability is zero except on an ellipse or at its center. This case is excluded if we require the distributions of x_1 and x_2 to be independent and Birnbaum, Raymond and Zuckerman [6] give the best possible result in this case as follows.

(xxxiv) The probability that (x_1, x_2) lies in

$$\frac{x_1^2}{s^2} + \frac{x_2^2}{t^2} \leqq 1$$

is not less than $1-P(s, t)$ where, assuming that $\sigma_1^2/s^2 \leqq \sigma_2^2/t^2$,

$$P(s, t) = 1 \quad \text{if} \quad \sigma_1^2/s^2 + \sigma_2^2/t^2 \geqq 1$$

$$P(s, t) = \frac{\sigma_1^2}{s^2} + \frac{\sigma_2^2}{t^2} - \frac{\sigma_1^2\left(1 - \frac{\sigma_1^2}{s^2} - \frac{\sigma_2^2}{t^2}\right)}{s^2 - \sigma_1^2}$$

if

$$\frac{\sigma_1^2}{s^2} + \frac{\sigma_2^2}{t^2} \leqq 1 \leqq \frac{1}{2}\left\{\frac{\sigma_1^2}{s^2} + \frac{2\sigma_2^2}{t^2} + \sqrt{\left(\frac{\sigma_1^4}{s^4} + \frac{4\sigma_2^4}{t^4}\right)}\right\},$$

$$P(s, t) = \frac{\sigma_1^2}{s^2} + \frac{\sigma_2^2}{t^2} - \frac{\sigma_1^2\sigma_2^2}{s^2t^2}$$

if

$$\frac{1}{2}\left\{\frac{\sigma_1^2}{s^2} + \frac{2\sigma_2^2}{t^2} + \sqrt{\left(\frac{\sigma_1^4}{s^4} + \frac{4\sigma_2^4}{t^4}\right)}\right\} \leqq 1$$

These are obtained with equality for distributions of x_1 and x_2 with only two or three values, and the proof is obtained by showing that the general case can be reduced to one of this kind

Chapelon called the ellipsoid $\mu_{20} \cdot \quad {}_0u_1^2 + 2\mu_{110} \cdots {}_0u_1u_2 + \cdots = 1$ a *quadrique type*, and the parallelepiped which circumscribes this and has faces parallel to the axial planes a *parallélépipède type*. He showed [10] that

(xxxv) the probability that $(x_1, \cdots, x_n)$ lies inside the region obtained by dilating the *quadrique type* or *parallélépipède type* about its center by a scale factor λ is not less than $1-n/\lambda^n$.

Leser [19] considered an ellipsoid with axes parallel to the coordinate planes and also imposed a monotonicity condition.

(xxxvi) Let P be the probability that

$$\sum\left(\frac{x_j}{\lambda_j\sigma_j}\right)^2 \leqq n, \qquad (\sigma_i^2 = E(x_i^2))$$

and let

$$n/\sigma_0^2 = \sum 1/\sigma_i^2, \qquad n/\lambda_0^2 = \sum 1/\lambda_i^2.$$

Let $A(R_0)$ be the average probability density in the space

$$\lambda_0^2\sigma_0^2 \sum \left(\frac{x_i}{\lambda_i\sigma_i}\right)^2 = R_0^2$$

and suppose that $A(R)$ is a non-increasing function of R for $R \leqq k\sigma_0\sqrt{n}$. Then if $k \leqq 1$ we have

$$P \geqq 0 \qquad \text{for} \quad \lambda_0 \leqq 1$$

$$P \geqq 1 - \lambda_0^{-2} \quad \text{for} \quad 1 \leqq \lambda_0:$$

if

$$1 \leqq k \leqq \sqrt{\left(1 + \frac{2}{n}\right)}$$

we have

$$P \geqq \frac{n+2}{2}\left(1 - \frac{1}{k^2}\right)\left(\frac{\lambda_0}{k}\right)^n \quad \text{for} \quad \lambda_0 \leqq \left(\frac{2}{n+2}\right)^{1/n} k,$$

$$P \geqq 1 - \frac{1}{k^2} \qquad \text{for} \quad \left(\frac{2}{n+2}\right)^{1/n} k \leqq \lambda_0 \leqq k$$

$$P \geqq 1 - \frac{1}{\lambda_0^2} \qquad \text{for} \quad k \leqq \lambda_0;$$

and if

$$\sqrt{\left(1 + \frac{2}{n}\right)} \leqq k$$

we have

$$P \geqq \left(\frac{n}{n+2}\right)^{n/2} \lambda_0^n \qquad \text{for} \quad \lambda_0 \leqq \left(\frac{2}{n+2}\right)^{1/n}\left(\frac{n+2}{n}\right)^{1/2},$$

$$P \geqq 1 - \left(\frac{2}{n+2}\right)^{2/n} \frac{1}{\lambda_0^2} \quad \text{for} \quad \left(\frac{2}{n+2}\right)^{1/n}\left(\frac{n+2}{n}\right)^{1/2} \leqq \lambda_0 \leqq \left(\frac{2}{n+2}\right)^{1/n} k,$$

$$P \geqq 1 - \frac{1}{k^2} \qquad \text{for} \quad \left(\frac{2}{n+2}\right)^{1/n} k \leqq \lambda_0 \leqq k,$$

$$P \geqq 1 - \frac{1}{\lambda_0^2} \qquad \text{for} \quad k \leqq \lambda_0.$$

Romanovski [29] gave an inequality concerning deviations of observed frequencies from true ones Let there be n mutually independent trials each with s mutually exclusive results and let p_i be the probability of the ith result and q_i the observed relative frequency. Then
(xxxvii)

$$Pr\left\{ \sum (q_i - p_i)^2 < \epsilon^2 \right\} > 1 - \left(1 - \frac{1}{s}\right) \Big/ n\epsilon^2,$$

while to compare two sets of trials (with the same p_i) we have that if n' and n'' trials give observed relative frequencies q_i' and q_i'' then

$$Pr\left\{ \sum (q_i' - q_i'')^2 < \epsilon^2 \right\} > 1 - \frac{1}{\epsilon^2}\left(1 - \frac{1}{s}\right)\left(\frac{1}{n'} + \frac{1}{n''}\right).$$

Camp [8] gave a best possible generalization of Tchebychef's inequality in terms of new statistics which he called contour moments. Let Q_λ be the set of points at which $f(x_1, \cdots, x_n) > \lambda$ and let x_λ be the measure of Q_λ Let the frequency function be bounded above by A so that $x_A = 0$ while $x_0 \leqq \infty$. From the monotonic decreasing function λ of x we obtain a function $y(x)$ which is defined for all x by making $y(x) = \min \lambda(x)$ if λ is multiple-valued at x, while $y(x)$ is the value of λ at the beginning of an interval in which λ is undefined. Let $\widehat{\mu}_r = \int_0^{x_0} x^r y(x) dx$. Then
(xxxviii) the probability that $(x_1, \cdots, x_n)$ belongs to Q_λ is not less than

$$1 - \frac{\widehat{\mu}_{2r}}{x_\lambda^{2r}} \left(\frac{2r}{2r+1}\right)^{2r}$$

A NOTE ON FUTURE DEVELOPMENTS

Although, as shown above, much work has been done, much still remains to do. From a purely mathematical point of view the possibilities for further work are endless, since the ideal situation is one in which there is a best possible inequality for every combination of data about the population. From the practical point of view the types of

data which need be considered are more limited, but as regards one-dimensional distributions it would be useful to have inequalities for any intervals in terms of any number of moments. The work of Selberg [33] is a step in this direction; a note in [35, p x], suggests that Markoff solved this problem but the solution does not seem to be extant. It would be interesting also to impose a restriction on the number of modes in the moment problem (thus continuing Royden's work in [30]), which would exclude the present solution for all sufficiently large n, or to find more inequalities depending on geometrical conditions For distributions in more than one dimension there are possibilities of generalizing almost all the results belonging to one dimension

A NOTE ON APPLICATIONS

The application of the inequalities listed in the preceding pages is, as indicated in the Introduction, to the estimation of the probability of observations falling in given ranges, in terms of some known properties of the distribution of the observations. Such problems arise in the setting of quality control limits, or in the devising of tests of hypotheses. Suppose, for example, that we want to test the hypothesis that the mean value of x is 0, and we know that its variance is 1. (Such a knowledge of variance could be derived from past experience, measurements on mass-produced items may have fairly stable variances, but their mean values may change owing to tool-wear.) From Tchebychef's inequality the probability that $|x|>10$, when the mean value of x is 0, is less than 1/100, so that if we reject the hypothesis if an observation is numerically larger than 10 we have a test with significance level 1 per cent. If we know that the distribution of x is unimodal then Selberg's result (xvii) gives us that the probability of $|x|>7.52\ldots$ is less than 1/100. If we know that the distribution of x is normal then the probability that $|x|>2.57\ldots$ is less than 1/100, from tables of the normal probability integral The more we know about the distribution the more we can contract the bounds on $|x|$. If we know that the variance is not greater than 1 then we can use Tchebychef's inequality as above; a more precise value of the variance (say .81) would lead to the narrower bounds $|x|>9$. Thus in place of exact knowledge of the moments used in the statements of theorems we can use inequalities satisfied by them

As regards geometrical conditions, which are difficult to verify, it may be worth remembering that in much statistical work we act on the principle that premises which are nearly correct lead to conclusions

which are nearly correct; for example, although no variate met with in practice can be exactly normally distributed, the normal distribution is often a sufficiently good approximation for tables of significance derived from it to be usefully employed. Similarly an inequality based on the assumption that a frequency function has only one maximum may reasonably be used when the frequency function is "nearly unimodal." What is an unsatisfactory state of affairs to the pure mathematician may be the only one available to the practicing statistician.

REFERENCES

The references which follow refer only to the most general result of each type and work which has been superseded is not mentioned A much fuller bibliography of the subject, which has been of the greatest assistance in compiling the present paper, was given by Savage [31]. I am also indebted to referees for a few references. Papers of which my knowledge is derived only from quotation elsewhere are marked*.

[1] Berge, P. O. "A note on a form of Tchebycheff's theorem for two variables," *Biometrika*, 29 (1937) 405–6.

[2] Bergström, H. "On the central limit theorem in the case of not equally distributed random variables," *Skandinavisk Aktuarietidskrift*, 32 (1949) 37–62.

*[3] Bernstein, S. "Sur une modification de l'inégalité de Tchebichef," *Ann. Sc. Instit Sav Ukraine*, Sect. Math. I (1924). (Russian, French Summary).

[4] Bernstein, S. "Sur quelques modifications de l'inégalité de Tchebycheff," *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS*, 17 (1937) 279–82

[5] Berry, A C "The accuracy of the Gaussian approximation to the sum of independent variates," *Transactions of the American Mathematical Society*, 49 (1941) 122–36.

[6] Birnbaum, Z. W , Raymond, J. and Zuckerman, H. S. "A generalization of Tshebyshev's inequality to two dimensions," *Annals of Mathematical Statistics*, 18 (1947) 70–9.

[7] Birnbaum, Z W "On random variables with comparable peakedness," *Annals of Mathematical Statistics*, 19 (1948) 76–81.

[8] Camp, B H "Generalization to N dimensions of inequalities of the Tchebycheff type," *Annals of Mathematical Statistics*, 19 (1948) 568–74.

*[9] Cantelli, F. P "Intorno ad un teorema fundamentale della teoria del rischio," *Bolletino dell' Associazione degl. Attuari Italiani*, Milan, (1910) 1–23

[10] Chapelon, J. M. "Sur l'inégalité fondamentale du calcul des probabilités," *Bulletin de la Société Mathématique de France*, Paris, 65 (1937) 100–8.

[11] Craig, C. C. "On the Tchebycheff inequality of Bernstein," *Annals of Mathematical Statistics*, 4 (1933) 94–102

[12] van Dantzig, D. "Une nouvelle généralisation de l'inégalité de Bienaymé," *Annales de l'Institut Henri Poincaré*, 12 (1951) 31–43.

[13] Guttman, L "An inequality for kurtosis," *Annals of Mathematical Statistics* 19 (1948) 277–8

[14] Guttman, L. "A distribution-free confidence interval for the mean," *Annals of Mathematical Statistics*, 19 (1948) 410–13.

[15] Hsu, P. L. "The approximate distributions of the mean and variance of a sample of independent variables," *Annals of Mathematical Statistics*, 16 (1945) 1–29.

[16] Khamis, S. H. "On the reduced moment problem," *Annals of Mathematical Statistics*, 25 (1954) 113–22.

[17] Kolmogoroff, A. N. "Über die Summen durch den Zufall bestimmter unabhängiger Grössen," *Mathematische Annalen*, 99 (1928) 309–19.

[18] Kolmogoroff, A. N. "Bemerkungen zu meiner Arbeit 'Über die Summen Zufalligen Grössen'," *Mathematische Annalen*, 102 (1929) 484–8.

[19] Leser, C. E. V. "Inequalities for multivariate frequency distributions," *Biometrika*, 32 (1942) 284–93.

[20] Lurquin, C. "Sur une inégalité fondamentale de probabilité," *Comptes Rendus*, Paris, 187 (1928) 868–70.

[21] Lurquin, C. "Sur un théoreme de limite pour la probabilité au sens de Bienaymé-Tchebycheff," *Bulletin Acad. Bruxelles*, (5) 14 (1928) 641–58.

[22] Meidell, B. "Sur la probabilité des erreurs," *Comptes Rendus*, Paris, 176 (1923) 280–2.

[23] von Mises, R. "Sur une inégalité pour les moments d'une distribution quasi-convexe," *Bulletin des Sciences Mathématiques* (2), 62 (1938) 68–71.

[24] von Mises, R. "The limits of a distribution function if two expected values are given," *Annals of Mathematical Statistics*, 10 (1939) 99–104.

[25] Narumi, S. "On further inequalities with possible application to problems in the theory of probability," *Biometrika*, 15 (1923) 245–53.

[26] Offord, A. C. "An inequality for sums of independent random variables," *Proceedings of the London Mathematical Society* (2), 48 (1945) 467–77.

[27] Pearson, K. "On generalized Tchebycheff theorems in the mathematical theory of statistics," *Biometrika*, 12 (1919) 284–96

[28] Robbins, H. "Some remarks on the inequality of Tchebychef," *Courant Anniversary Volume*, New York: Interscience Publishers, Inc., (1948) 345–50.

*[29] Romanovski, V. I. "On inductive conclusions in statistics," *Comptes Rendus (Doklady) de l'Académie des Sciences de l'URSS* (N.S.), 27 (1940) 419–21.

[30] Royden, H. L. "Bounds on a distribution function when its first *n* moments are given," *Annals of Mathematical Statistics*, 24 (1953) 361–76.

[31] Savage, I. R. "Bibliography of Nonparametric Statistics and Related Topics," *Journal of the American Statistical Association*, 48 (1953) 844–906.

[32] Selberg, H. L. "Über eine Ungleichung der Mathematischen Statistik," *Skandinavisk Aktuarietidskrift*, 23 (1940) 114–20.

[33] Selberg, H. L. "Zwei Ungleichungen zur Ergänzung des Tchebycheffschen Lemmas," *Skandinavisk Aktuarietidskrift*, 23 (1940) 121–5.

[34] Selberg, H. L. "On an inequality in mathematical statistics," *Norsk Matematisk Tidsskrift*, 24 (1942) 1–12, (Norwegian)

[35] Shohat, J. A. and Tamarkin, J. D. "The problem of moments," *American Mathematical Society*, New York, 1943.

[36] Smith, C. D. "On generalized Tchebycheff inequalities in Mathematical Statistics," *American Journal of Mathematics*, 52 (1930) 109–26.

[37] Smith, C. D. "On Tchebycheff approximation for decreasing functions," *Annals of Mathematical Statistics*, 10 (1939) 190–2.

[38] Wald, A. "Limits of a distribution function determined by absolute moments and inequalities satisfied by absolute moments," *Transactions of the American Mathematical Society*, 46 (1939) 280–306.

[39] Winsten, C. B. "Inequalities in terms of mean range," *Biometrika*, 33 (1946) 283–95.

[40] Zelen, M. "Bounds on a distribution function that are functions of moments to order four," *Journal of Research of the National Bureau of Standards*, 53 (1954) 377–81.

THE RANDOMIZATION THEORY OF EXPERIMENTAL INFERENCE*

Oscar Kempthorne
Iowa State College

The paper contains a description of the extent to which the use of randomization in experimental designs permits evaluation of the experimental results. The case considered is that in which the whole population of treatments is used with particular experimental material. A completely general mathematical specification of the design is given and the procedure by which linear models for the experimental results are derived is exemplified by the cases of the completely randomized design, randomized blocks, Latin squares, and a particular systematic design. The case of the completely randomized design is discussed extensively. An assessment of the present state of randomization theory is given, with a statement of major deficiencies.

INTRODUCTION

It has become general in experimentation to insist on the physical act of randomization, which in all cases is randomization over a finite universe of possibilities. Frequently the analysis is then presented by way of least squares as regards estimation and general linear hypothesis theory, as regards tests of significance, the construction of intervals and so on. For example, Yates [22] has given a nominally exact test for treatment differences when data are missing from the planned randomized experiment This test is based on the assumption of a linear model with errors which are normally and independently distributed around zero with constant variance. That this assumption is not validated by randomization is clear: for example, in the case of randomized blocks there is a negative correlation between the observed values for two treatments in any one block A treatment observation in a block may be high because the treatment fell on the highest yielding plot, and if a treatment so falls then another treatment will fall on a lower yielding plot and hence give a lower value.

Again, it has become general to insist that a design satisfy the property of unbiasedness [25], whereas if the observations satisfy the conditions of the simple Markoff theorem, or the general linear hypothesis this property is automatically satisfied. Designs otherwise entirely

* Journal paper number J-2675 of the Iowa Agricultural Experiment Station, Ames, Iowa. Project 890. This paper is based on one presented to the American Statistical Association at Chicago on December 29, 1952

valid from the point of view of least squares have been removed from the catalogue of reasonable designs, because the condition of unbiasedness is not satisfied, for example the semi-Latin square [24].

Tests of significance in the randomized experiment have frequently been presented by way of normal law theory, whereas their validity stems from randomization theory. The use of tests of significance based on randomization of the observations was given by Fisher in 1935 in the first edition of *The Design of Experiments* [21]. Here the problem was the comparison of two paired samples of 15 observations. After commenting on the tendency of theoretical statisticians "to stress the normality assumption," Fisher states that "it seems to have escaped recognition that the physical act of randomization . . , affords the means, in respect of any particular body of data, of examining the wider hypothesis in which no normality of distribution is implied" In the particular case with the two samples of 15 observations the actual difference between the totals is 314. Fisher then notes that in just 1726 of the 2^{15} possible partitions of the 15 pairs of observations into two paired samples of 15, a difference as large as or larger than that observed would be obtained. Hence a significance level of 5 267 per cent can be attached to the null hypothesis that there is no difference between the two populations For comparison Fisher notes that the normal theory test gave a significance level of 4.97 per cent and when Yates's correction for continuity is used, gave a significance level of 5 158 per cent. In a later paper, Fisher [11], after describing the randomization method states ". . . conclusions have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method "

Even though the application of general linear hypothesis theory appears in many cases to lead to inferences which are essentially correct by Fisher's criterion above, the validity of such normal theory inferences in the case of some designs, notably small Latin squares, is highly questionable. In the case of the 3×3 Latin square, the only randomization test of the null hypothesis is of size 50 per cent whereas normal theory allows any size of test. Some other confusions existing in the literature are:

(1) the lack of distinction made between the analysis of randomized blocks and the two-way classification with normally distributed errors, i.e , data assumed to follow the normal law model

$$y_{ijk} = \mu + a_i + b_j + e_{ijk}, \qquad i = 1, 2, \cdots, r;$$
$$j = 1, 2, \cdots, s; k = 0, 1, 2, \cdots, n_{ij}$$

where μ, a_i and b_j are constants and the e_{ijk} are normally and independently distributed around zero with common variance. From another point of view the randomized block experiment is not symmetrical with respect to blocks and treatments, whereas the above model is symmetrical with respect to the a_i and the b_j.

(2) the question of the universe for which the conclusions of the randomized experiment hold. More fully, is the inference from the randomized block experiment valid only for the population of experimental units sampled at random, or for some broader (always unspecified) population of experimental units. It appears that the former is the case.

(3) the respective natures of the inferences in the randomized block experiment without and with the analysis of covariance. The analysis of covariance gives apparent increased precision of estimates, over what is obtained with randomized blocks, but the two analyses, without and with covariance, are not of equal validity, the latter utilizing more assumptions

(4) the basis for evaluation of the efficiency of one randomized design relative to another.

All the methods under discussion herein were originated primarily by Fisher, linear model theory [6], extended as the fitting of constants by Yates [23], the use of randomization [8], the use of randomization tests [9], the analysis of covariance [7].

It is of interest to note in the past few years some explicit recognition of the situation, for example by Grundy and Healy [13], Barnard [2], and Tukey [20]. The author's book [14] contains descriptions of both normal law analyses and randomization analyses for the basic designs. However, a concise connected statement of the whole situation is needed with an account in general terms of randomization analyses.

SCOPE OF THE PAPER

This paper will deal with only one type of experiment, namely, the comparative experiment in which the experimenter wishes to make statements about the differences in response or yield produced by a fixed set of treatments. Other comparative experiments are of the type in which there is a population of experimental units and a population of treatments, but only random samples taken in specified ways of units and treatments are used. The basic considerations to be given apply to broader classes of experiment, but the limitation to the stated type is necessary in order to give a reasonably complete account. The

absolute experiment as defined by Anscombe [1] is not discussed. The aim of the paper is to show how randomization gives a fairly complete basis for statistical inference from the comparative experiment, particularly if additivity as defined later holds, and to exhibit the gaps in our present knowledge.

CRITERIA OF VALIDITY OF EXPERIMENTAL INFERENCE

The paper by Anscombe [1] contains a detailed discussion of the criteria of validity of inferences, but with special emphasis on tests. The crucial aspects of a statistical experimental inference in comparative experiments are that we wish to estimate treatment differences and comparisons and, because a point estimate is of little utility without a measure of its reliability, we must have a measure of reliability. In order to use the measure of reliability with confidence we should know the joint distribution of the estimate and its estimated reliability. Finally as a guide to the future actions of the experimenter, we need accurate tests of significance.

The important aspect of the validity of an inference is the validity of the assumptions on which the inference is based. In simple experimental situations one can assume a model which specifies that the observation is equal to its expectation plus an error which is normally and independently distributed with constant variance and then apply standard linear hypothesis theory. Such a procedure can be questioned because the amount of data necessary to verify the assumptions is usually outside the ability of the experimenter to collect. The making of assumptions of normality and then applying the battery of mathematical statistical tests is not a satisfactory basis for experimental inference, because the extent to which the reliability of an inference depends on the assumptions made in the analysis is usually unknown.

The great contribution of Fisher to experimental inference was the introduction of randomization [8] together with the later recognition that randomization tests could be made.[1] It is rather curious that the introduction of randomization should permit the making of more accurate inferences than are possible without its use, more accurate in the sense that the probability statements that are made are more accurate. The inferences also have a definite relation to what is conceptually observable in the situation.

There have been attempts to base experimental inference on the

[1] The author believes that modern probability sampling also was introduced by Fisher and his co-workers at Rothamsted in the late 1920's. In this case randomization is the crucial element distinguishing modern sampling from that done early in this century.

fitting of trend lines, for example, orthogonal polynomials, or autoregressive models as well as treatment parameters. These are essentially the same nature as the use of the anlysis of covariance, which suffers from the defect that an assumption relating the response to the concomitant variable is necessary. It is not surprising that an increase in the assumptions results in apparent increased accuracy of estimates, but it may be debated whether this increased apparent accuracy is really true accuracy.

DESCRIPTION OF MATHEMATICAL TREATMENT

Consider a class C of possible patterns for application of treatments to experimental units, such as for example the 12 possible 3×3 Latin squares, with particular members denoted by c An element c is chosen from the class C according to some probability distribution Let us suppose that the experimental units are denoted by u which runs from 1 to N, and the treatments by j which runs from 1 to t A particular experimental pattern is specified by N pairs of numbers (uj), u denoting the experimental unit and j the treatment imposed on that unit. Let

$$\delta_{1\,2\,\cdots N}^{j_1 j_2 \cdots j_N}$$

be unity if treatment j_1 is imposed on the first experimental unit, treatment j_2 on the second and so on, and be zero otherwise Then the probability distribution according to which c is chosen from C is given by the numbers

$$P(\delta_{1\,2\,\cdots N}^{j_1 j_2 \cdots j_N} = 1).$$

We are at liberty to choose any set of t^N nonnegative numbers whose sum is unity for the numbers

$$P(\delta_{1\,2\,\cdots N}^{j_1 j_2 \cdots j_N} = 1),$$

but should consider the consequences of each particular choice. We also need certain marginal properties of the distribution of the

$$\delta_{1\,2\,\cdots N}^{j_1 j_2 \cdots j_N}$$

We will need, for example, to talk about $\delta_u{}^j$, which equals unity if treatment j is applied to experimental unit u and zero and otherwise, and about $P(\delta_u{}^j=1)$. Or for example we may need numbers such as $P(\delta_v{}^k=1 \mid \delta_u{}^j=1)$, which is the probability that treatment k is applied to experimental unit v, given that treatment j is applied to

unit u. The actual realization of a particular pattern is accomplished by the use of a random device or random numbers. A basic assumption is that we have a random device which generates a distribution with the stated properties.

It is possible by this means to represent all standard types of experimental design in a unified way, that is, systematic, completely randomized designs, complete randomized blocks and Latin squares, split-plot designs and crossover designs, as well as other types which can be visualized but have never, as far as I know, been used.

The general procedure is to examine the behavior of certain functions of the observations obtained from each pattern. The experimental units will be represented by suffices which indicate the amount of stratification in the design For example, in the randomized block experiment the units will be denoted by (uv), where u gives the block number and v the number of the plot within the block. It is frequently convenient to use an identity such as the following, which is useful with randomized blocks

$$x_{uv} = x_{..} + (x_{u.} - x_{..}) + (x_{uv} - x_{u.}) \tag{1}$$

where u runs from 1 to r, v runs from 1 to s and

$$x_{u.} = \frac{1}{s}\sum_{v} x_{uv}, \qquad x_{..} = \frac{1}{rs}\sum_{uv} x_{uv}.$$

This identity is useful in the particular case because it will be seen (15) that $x_{..}$, $x_{u.}-x_{..}$ appear as constants while $(x_{uv}-x_{u.})$ is multiplied by a random variable. The general procedure will be exemplified for the cases of the completely randomized design, the randomized block design and the Latin square.

The delta quantities used here play the same role as when they are used in finite sampling theory They are the imposed random variables in the experimental design and it is therefore reasonable that they should be exhibited explicitly. Also it turns out that they enable succinct presentation of the distributional problems and of the solution of these problems

ADDITIVITY

The term "additivity" has been used rather freely in discussions of experimental designs, without precise meaning being attached to it. For instance a model

$$y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk}$$

is an additive model in the sense that the right hand side is the sum of a number of parameters as opposed to product or a mixture of sums and products. Also if the symbols i, j are used to denote levels of factors, it is usually said that additivity of the factors holds if the terms $(ab)_{ij}$ are unnecessary.

A somewhat different concept of additivity is used in randomization analysis. We suppose that each experimental unit would have a response or yield, say x_u, for the u^{th} unit under some basic condition. Then additivity implies that the yield of the u^{th} unit under treatment j which we may denote by y_{uj} is given by

$$y_{uj} = x_u + t_j. \tag{2}$$

Consequences of this are, for example,

$$y_{uj} - y_{uj'} = t_j - t_{j'},$$

which states that if we could apply both treatments j and j' to each experimental unit and observe the difference, then this difference would be constant over the experimental units.

The concept we have given above may be termed additivity in the strict sense; if instead of (1) we have

$$y_{uj} = x_u + t_j + \epsilon_{uj} \tag{3}$$

where the ϵ_{uj}'s are independent random variables over all u and j, we can say that we have additivity in the broad sense. The estimation procedures and tests are identical in the two cases, but the distinction is worth making because from it we obtain indications of when randomization is desirable and when it is probably unimportant. Additivity in the present context will therefore be used to mean absence of interaction of treatment and ultimate experimental unit. It is different from additivity as used in the previously mentioned cases because in those cases the separate identity of treatments and experimental units is not maintained. In the case when the units are arranged in groups it is desirable to weaken this assumption to one of additivity of treatments within groups but not between groups.

Linear models for the case when the additivity assumption is weakened have been examined by Wilk (unpublished manuscript).

THE COMPLETELY RANDOMIZED DESIGN

In this case with s treatment we have rs equals N experimental units and each pattern in which every treatment is represented r times

is given equal probability. If δ_u^j is unity when treatment j occurs on plot u and zero otherwise, then

$$P(\delta_u^j = 1) = \frac{1}{s}, \qquad P(\delta_u^j = 0) = \frac{s-1}{s}$$

$$P(\delta_{u'}^j = 1 \mid \delta_u^j = 1) = \frac{r-1}{rs-1}, \quad P(\delta_u^{j'} = 1 \mid \delta_u^j = 1) = 0 \tag{4}$$

$$P(\delta_{u'}^{j'} = 1 \mid \delta_u^j = 0) = \frac{r}{rs-1},$$

and so on.

Under additivity the yield on plot u, with treatment k is

$$\begin{aligned} y_{uk} &= x_u + t_k \\ &= x_. + t_k + (x_u - x_.). \end{aligned}$$

The observed total yield for treatment k is equal to the sum of the y_{uk}'s for the plots on which treatment k falls, so that we may write

$$T_k = r(x_. + t_k) + \sum_{u=1}^{n} \delta_u^k (x_u - x_.) \tag{5}$$

or

$$\frac{T_k}{r} = x_. + t_k + \frac{1}{r} \sum_{u=1}^{n} \delta_u^k (x_u - x_.). \tag{6}$$

In equation (6), the observed mean yield is expressed as the sum of the true mean yield with treatment k, $(x + t_k)$, plus an error

$$\frac{1}{r} \sum_{u=1}^{n} \delta_u^k (x_u - x\,).$$

This error is a linear function with coefficients $(x_u - x_.)/r$ of the random variables in the experiment, the δ_u^k. We know the distributional properties of these random variables. Hence we may find the distributional properties of the treatment means and any function of them, and of the individual observed yields and the sum of squares of yields within treatments.

Writing the observations in a form such as (6) has very strong advantages in that it exhibits precisely the nature of the error of an observation and expresses this error in terms of elementary random variables whose distributional properties are known. This simplifies con-

siderably the problems of examining distributional properties of the error, that is of the quantities

$$\sum_{u=1}^{n} \delta_u^k \frac{(x_u - x_.)}{r}$$

for the possible values of k The quantities $(x_u - x_.)$ are not to be regarded as random variables, but as fixed unknown quantities which are attached to the treatment observations according to a known distribution. The process of taking expectations is simplified by noting that δ_u^k; is either zero or unity. Hence

$$E[(\delta_u^k)^r] = P(\delta_u^k = 1),$$

for all non-zero r and

$$E[(\delta_u^k)^p(\delta_{u'}^{k'})^q] = P(\delta_u^k = 1,\ \delta_{u'}^{k'} = 1),$$

for all non-zero p and q

It is easily seen that a treatment mean is an unbiased estimate of $(x_. + t_k)$, for

$$E\left[\frac{T_k}{r} - (x_. + t_k)\right] = E\sum_{u=1}^{n} \delta_u^k \frac{(x_u - x_.)}{r}$$

$$= \frac{1}{rs}\sum_{u=1}^{n}(x_u - x_.) = 0. \tag{7}$$

Also the variance of a treatment mean is

$$E\left[\frac{T_k}{r} - (x_. + t_k)\right]^2$$

$$= \frac{1}{r^2} E\left\{\sum_u \delta_u^k (x_u - x_.)\right\}^2$$

$$= \frac{1}{r^2} E\left\{\sum_u (\delta_u^k)^2 (x_u - x_.)^2 + \sum_{u' \neq u} \delta_u^k \delta_{u'}^k (x_u - x_.)(x_{u'} - x_.)\right\}$$

$$= \frac{1}{r^2}\left\{\sum_u \frac{1}{s}(x_u - x_.)^2 + \sum_{u' \neq u} \frac{r-1}{s(rs-1)}(x_u - x_.)(x_{u'} - x_.)\right\}$$

$$= \frac{s-1}{rs(rs-1)}\sum_u (x_u - x_.)^2$$

$$= \left(\frac{s-1}{s}\right)\frac{1}{r}\sigma^2, \quad \text{where} \quad \sigma^2 = \frac{1}{rs-1}\sum_u (x_u - x_.)^2. \tag{8}$$

Also, if A and B are two contrasts of the observed treatment means,

$$A = \sum_j \lambda_j \frac{T_j}{r}, \qquad \sum_j \lambda_j = 0,$$

$$B = \sum_j \nu_j \frac{T_j}{r}, \qquad \sum_j \nu_j = 0,$$

then it may be shown that

$$\mathrm{Cov}\,(A, B) = \left(\sum_j \lambda_j \nu_j\right)\frac{\sigma^2}{r}, \tag{9}$$

and thence that

$$V(A) = \left(\sum_j \lambda_j^2\right)\frac{\sigma^2}{r}, \qquad V(B) = \left(\sum_j \nu_j^2\right)\frac{\sigma^2}{r}\cdot \tag{10}$$

The covariance of two treatment means is therefore $-(\sigma^2/rs)$, which is the usual covariance in a finite sampling situation. Furthermore, we may make the following subdivision of the total sum of squares among the observations and find their expectations

	Sum of Squares	*Expectation of Sum of Squares*
Between treatments	$\frac{1}{r}\sum_k T_k^2 - \frac{1}{rs}(\sum T_k)^2$	$(s-1)\sigma^2 + r\sum_k (t_k - \bar{t})^2$
Within treatments	*by difference*	$s(r-1)\sigma^2$
Total	$\sum$ (observations)$^2 - \frac{1}{rs}(\sum T_k)^2$	

If we divide the sums of squares by $(s-1)$ and $s(r-1)$ respectively we have mean squares both of whose expectations are σ^2, if the t_k's are equal It is natural to call the quantities $(s-1)$ and $s(r-1)$ the degrees of freedom, because the sums of squares can be expressed as the sum of squares of the respective number of random variables, which are uncorrelated and have the same variance, σ^2.

The analysis of variance given above could be obtained by the use of the linear model

$$y_{uk} = \mu + t_k + \epsilon_{uk}, \quad u = 1, 2, \cdots, r;\ k = 1, 2, \cdots, s \tag{11}$$

where y_{uk} is the yield of the kth treatment on the uth plot on which it

occurs, μ and t_k are constants and the ϵ_{uk} are uncorrelated errors with expectation of zero and the same variance. To establish the property of unbiasedness for this design it is (by definition, [25]) necessary to show that the expectation over randomizations of the error mean square resulting from this model is equal to the mean square among all observations in the absence of treatment effects The property of unbiasedness then determines whether the error mean square obtained by the use of least squares has an expectation over randomizations equal to the true randomization error mean square in the absence of treatment effects. It therefore relates to the validity of the analysis of the model by simple least squares, this validity being ascertained by comparison with randomization analysis. In passing it should perhaps be noted that the property has no intrinsic relation to the concept of unbiasedness of a test.

We have shown so far that the use of randomization and additivity together give us essentially the same results for estimates as we obtain by the use of the normal theory model in which the ϵ_{uk}'s of equation (11) are also normally distributed. However, we may note the following differences:

(1) With randomization and additivity, as we have defined it, each observation is automatically subject to the same variance, whereas this must be assumed in the application of (11);

(2) A treatment mean has a variance of $[(s-1)/s]\ \sigma^2/r$, but *may be regarded* as having a variance of σ^2/r for comparison with other treatment means in the same experiment.

We should note that the main results we have given above are essentially true also when there is additivity in the broad sense, that is, there is an error additional to the error $\sum_u \delta_u^k(x_u - x_.)$. This latter error may be conveniently termed the plot error, and the total error consists of the plot error plus the additional error (which may be, for example, a measurement error).

Randomization affects only the plot error, and this statement indicates the fields of application where randomization is important. One can envisage experimental situations where plot errors are trivial and the additional errors large in comparison to them. In such a situation the lack of randomization will not seriously invalidate the experimental conclusions. This seems to be the main reason why the randomized experiment has not been essential to progress in some of the physical and chemical sciences.

There remain the questions of the joint distribution of the estimates and their estimated variances and of tests of hypotheses about the existence of treatment effects. The distributional problem is in principle straightforward, although its exact solution may be unattainable. Tests of hypotheses may be tackled by the randomization test procedure. We choose a function of the observations which we evaluate for the experimental pattern used and for each of the possible patterns with probabilities specified by the randomization procedure followed.

This latter will give rise to a discrete distribution of the function and our test procedure is to note the proportion of cases weighted by their probabilities in which the value of the function exceeds the value in the pattern which was used. If this proportion is equal to p per cent, then we can state that the significance level to be attached to the null hypothesis that treatments have no effect is equal to p per cent. If we use a 5 per cent test, we shall reject the null hypothesis if p is less than 5 per cent.

We should note here that the test procedure is described in such a way that we may delimit the set of possible configurations we shall use in any way we like, and can give whatever probability we like to particular plans, as far as the full randomization test procedure is concerned. If we have 20 experimental units which are in a row and are comparing 4 treatments and hence have 5 replicates, we could give zero probability to the patterns in which the treatments occur along the row in consecutive groups of five units. Such a procedure presents difficulties with regard to estimation, however, for the analysis of variance is intimately related to equal probabilities for all possible arrangements, and as far as I know the estimation procedure corresponding to any joint distribution of the $\delta_u{}^i$'s, other than those for the basic designs, has not been worked out though this might be a useful as well as amusing problem. Off-hand the balanced relations of the basic designs appear to be crucial for a simple analysis of experiment.

We now consider an approximation to the randomization test procedure for the completely randomized design, this approximation being suggested by the normal law theory. We note that under the null hypothesis, the treatment sum of squares plus the error sum of squares is equal to a constant, namely,

$$\sum_u (x_u - x_.)^2, \quad \text{or} \quad (rs - 1)\sigma^2,$$

which may be conveniently denoted by S_2. The normal theory test would be to compare

$$\left(\frac{\text{treatment mean square}}{\text{error mean square}}\right)$$

with the F distribution Since the sum of squares for treatments and sum of squares for error add to a constant, it is appropriate to consider the distribution over randomizations of the treatment sum of squares T divided by the total sum of squares S_2, i e , T/S_2

We have already seen that under the null hypothesis,

$$E(T) = \frac{(s-1)}{(rs-1)} S_2.$$

The variance of T is a rather complicated expression, but as r gets large, it is approximately equal to $2(s-1)/r^2s^2\ S_2^2$. The expectation of T/S_2 is $(s-1)/(rs-1)$, which is the mean of the beta distribution corresponding to the normal theory F test The variance of the beta distribution with parameters $(s-1)$ and $(rs-1)$ is

$$\frac{2(r-1)s(s-1)}{(rs-1)^2(rs+1)}.$$

As r gets large this is closely approximated by $2(s-1)/r^2s^2$, which is the value obtained for the randomization distribution as r gets large. Insofar as the randomization distribution of T can be represented by the first two moments, this distribution will be closely represented by the corresponding beta distribution if r is large. In a few cases that I have examined the approximation works well when r is small, for example, 4. It is, therefore, plausible that the randomization test can be approximated by the normal theory analysis of variance test. This serves as some theoretical basis for the fact which has been noticed by most statisticians, that the level of significance of the analysis of variance test for differences between treatments is little affected by the choice of a scale of measurement for analysis.

COMPLETE RANDOMIZED BLOCKS

This case will be reviewed rather briefly because detailed treatment is given in [14], Section 8 2. Here, only the case when additivity in the strict sense holds will be considered. As before the lessening of assumptions to additivity in the broad sense results in little change.

Denote the conceptual yield with treatment $k(=1, 2, \cdots, s)$ on plot $v(=1, 2, \cdots, s)$ of block $u(=1, 2, \cdots, r)$ by y_{uvk}. Under additivity

$$y_{uvk} = x_{uv} + t_k. \tag{12}$$

Now

$$y_{uvk} = y_{...} + (y_{u..} - y_{...}) + (y_{uvk} - y_{uv.}) + (y_{uv.} - y_{u..}) \tag{13}$$

or, using (12)

$$y_{uvk} = x_{..} + t_. + (x_{u.} - x_{..}) + t_k - t_. + (x_{uv} - x_{u.}). \tag{14}$$

We observe the yield from treatment k on a random plot v of each block u The observed yield we write as z_{uk} and we can write

$$z_{uk} = \mu + b_u + t_k + \sum_v \delta_{uv}{}^k (x_{uv} - x_{u.}) \tag{15}$$

where

$\mu = x_{..} + t_.$, $\quad b_u = x_{u.} - x_{..}$, $\quad$ and t_k is written for $(t_k - t_.)$,

or as

$$z_{uk} = \mu + b_u + t_k + \epsilon_{uk}, \tag{16}$$

which is the usual form for the model, with an obvious correspondence of symbols The distributional properties of $\delta_{uv}{}^k$ for the complete randomized block design are easily written out Hence, theoretically at least, the distributional properties of any function of the observations can be obtained The results on the estimates of treatment differences and their errors are with one exception the results on estimation obtained with normal law theory. The exception, which occurred also in the previous case of completely randomized designs, is that treatment totals or means are negatively correlated in the usual way for samples from finite populations. This correlation is such that we may for the purposes of comparing treatments within the experiment attribute a variance of σ^2/r to each treatment mean. However, the variance of a treatment mean for comparison with outside data is $[(s-1)/s]\sigma^2/r$ and not σ^2/r which is the formula generally used.

It was shown by Welch [21] and Pitman [17] that the randomization test may be approximated by the normal theory F test with $(s-1)$ and $(r-1)(s-1)$ degrees of freedom. Hence under some circumstances the randomization distribution is probably represented fairly accurately by the corresponding normal theory distribution, so that the standard F test may be used Some empirical support of this statement was obtained earlier by Eden and Yates [5].

An interesting facet of the randomization situation is the distribution

of the error sum of squares. It appears that the variance of the error sum of squares is

$$\frac{2(r-1)(s-1)}{r}\sigma^4$$

over the randomization distribution compared to $2(r-1)(s-1)\sigma^4$ over the normal theory distribution.

THE EFFICIENCY OF RANDOMIZED BLOCKS

The accuracy of Yates's [24] method of estimating the efficiency of randomized blocks relative to the completely randomized design has been questioned [4]. This is not surprising since Yates presents the analysis in terms of a normal law model, while the efficiency is based on a randomization argument. Yates's procedure is as follows, given the analysis of variance in the table below for the observed yields.

	d.f.	*Mean Square*
Blocks	$(r-1)$	B
Treatments	$(s-1)$	T
Error	$(r-1)(s-1)$	E

Consider the error mean square in the absence of treatment effects if blocks had not been used. This is obtained by substituting E for T and getting the overall mean square, which is

$$\frac{(r-1)B + r(s-1)E}{rs-1}.$$

This divided by E is the estimated efficiency of the blocks.

This is derived easily by the use of the previous notation. Number the plots by u and v, the block number and the number of the plot within the block respectively. The difference between the two designs is with respect to the distribution of the quantities δ_{uv}^k. In the case of the completely randomized design

$$P(\delta_{uv}^k = 1) = \frac{1}{t}$$

as before, but

$$P(\delta_{uv'}^k = 1 \mid \delta_{uv}^k = 1) = \frac{(r-1)}{rs-1}, \qquad v' \neq v,$$

compared to zero for randomized blocks. This is the only difference in the distributions which is needed.

The yield y_{uvk} satisfied

$$y_{uvk} = x_{uv} + t_k$$
$$= x_{..} + t_k + (x_{uv} - x_{..}).$$

So the observed yields z_{uk} in the completely randomized design satisfy or $z_{uk} = \mu + t_k + \sum_{uv} \delta_{uv}^{k}(x_{uv} - x\ )$. As we have seen the error mean square has an expectation of

$$\frac{1}{(rs-1)} \sum_{uv} (x_{uv} - x_{.}\ )^2 = \sigma_{CRB}^2, \quad \text{say.}$$

However,

$$\sum_{uv} (x_{uv} - x_{..})^2 = s \sum (x_u\ - x_{.})^2 + \sum_{uv} (x_{uv} - x_u\ )^2$$

or

$$(rs - 1)\sigma_{CRB}^2 = (r-1)B + r(s-1)\sigma_{RB}^2.$$

The estimation procedure follows directly from this equation since the expectation of the error mean square equals σ_{RB}^2.

MODIFICATIONS OF RANDOMIZED BLOCKS

Incomplete randomized blocks are from the present point of view rather minor modifications of complete randomized blocks Split-plot designs have a different joint distribution of the δ's Suppose the notation is as follows:

block number	$u = 1, 2, \cdots, r$
whole-plot number	$v = 1, 2, \cdots, t$
split-plot number	$w = 1, 2, \cdots, s$
whole-plot treatment	$j = 1, 2, \cdots, t$
split-plot treatment	$m = 1, 2, \quad \cdots, s.$

Then, for example,

$$P(\delta_{uvw}^{jm} = 1) = \frac{1}{ts}$$

$$P(\delta_{uvw'}^{jm'} = 1 \mid \delta_{uvw}^{jm} = 1) = \frac{1}{s-1}, \qquad w' \neq w,\ m' \neq m$$

and so on. The analysis is developed from this point of view in ([14], Section 19-1).

THE LATIN SQUARE WITH ADDITIVITY

Let y_{uvk} be the yield for treatment $k(=1, 2, \cdots, t)$ on the plot in row $u(=1, 2, \cdots, t)$ and column $v(=1, 2, \cdots, t)$. Then we may write for the observed total T_k for treatment k

$$T_k = t(\mu + r_u + c_v + t_k) + \sum_{uv} \delta_{uv}{}^k e_{uv} \tag{17}$$

where

$$e_{uv} = (x_{uv} - x_{u.} - x_{.v} + x_{..}). \tag{18}$$

If we take

$$\begin{aligned} P(\delta_{uv}{}^k = 1) &= \frac{1}{t} \\ P(\delta_{u'v}{}^k = 1 \mid \delta_{uv}{}^k = 1) &= 0, \quad u' \neq u, \\ P(\delta_{uv'}{}^k = 1 \mid \delta_{uv}{}^k = 1) &= 0, \quad v' \neq 0, \end{aligned} \tag{19}$$

and

$$P(\delta_{u'v'}{}^k = 1 \mid \delta_{uv}{}^k = 1) = \frac{1}{t-1}, \quad u' \neq u,\ v' \neq v.$$

and denote $[1/(t-1)^2]\sum_{uv} e_{uv}{}^2$ by σ^2, we find, for $\sum \lambda_k = 0$, $\sum \nu_k = 0$, that

$$E\left(\sum \lambda_k \frac{T_k}{t}\right) = \sum \lambda_k t_k,$$

$$\operatorname{Cov}\left(\sum \lambda_k \frac{T_k}{t}, \sum \nu_k \frac{T_k}{t}\right) = \left(\sum \lambda_k \nu_k\right) \frac{\sigma^2}{t}.$$

The expectation of the error mean square is σ^2, so that the elementary normal law procedures are obtained. The conditions given in (19) are mathematical expressions of conditions given by Fisher in his fundamental paper on the subject [8].

The normal approximation to the randomization test is not as easy to examine as in the previous cases. The quantity U

$$U = \frac{\text{treatment sum of squares}}{\text{error sum of squares plus treatment sum of squares}}$$

is easily seen to have an expectation of $1/(t-1)$, by the argument given given by Fisher [10]. Welch [21] found that the variance of U, depended on the population of Latin Squares from which the random one was chosen, in fact, on the transformation set, and on functions of the e_{uv}'s. Welch found indications that the significance level indicated by

the F table could be in error appreciably. With some constructed data, he estimated that the 5 per cent tabular F value would be exceeded in a proportion of 2.7 to 6.4 per cent of randomizations. It is interesting to speculate whether there exists a large family of Latin Squares different from the families specified by Fisher [8] and by Fisher and Yates [12], which has better properties with respect to the approximation of the F distribution to the randomization distribution of the analysis of variance criterion.

THE KNUT VIK SQUARE

It is not my aim here to consider in general the Knut Vik Square which is the classic case of a systematic design, but it seems desirable to include at least one example of a systematic design to show the generality of the approach The Knut Vik Square is as follows:

$$\begin{matrix} A & B & C & D & E \\ D & E & A & B & C \\ B & C & D & E & A \\ E & A & B & C & D \\ C & D & E & A & B \end{matrix}$$

If rows are characterized as levels 0, 1, 2, 3 and 4 of a factor x, and columns as levels 0, 1, 2, 3, 4 of a factor y, the 24 degrees of freedom between the 25 cells can be represented by 6 sets of 4 degrees of freedom, which are usually denoted by X, Y, XY, XY^2, XY^3 and XY^4. The partitions corresponding to XY and XY^4 are diagonal partitions, XY^2 corresponds to the above square and XY^3 to another Knut Vik square. It has been suggested [16] that the experimenter should choose XY^2 or XY^3 to give the treatment pattern and then use the other for the estimation of error If we number the rows by $u(=0, 1, 2, 3, 4)$, the columns by $v(=0, 1, 2, 3, 4)$ and treatments by $i(=0, 1, 2, 3, 4)$, then under random assignment of treatments to letters, a random choice of the degrees of freedom XY^2 or XY^3 to be associated with treatments, the observed treatment total becomes

$$T_k = t(\mu + \rho_u + \gamma_v + t_k) + \sum_{uv} \delta_{uv}{}^i(x_{uv} - x_{u.} - x_{.v} + x_{..})$$

where the ρ's, γ's and t's are row, column and treatment contributions, and $\delta_{uv}{}^i$ equals unity if treatment i appears in plot u and is zero otherwise. In this case with the Latin Square

$$P(\delta_{uv}{}^i = 1) \doteq \tfrac{1}{5}$$

but $P(\delta_{u'v'}{}^i = 1 \mid \delta_{uv}{}^i = 1)$ is not a constant for all $(u'v')$ unequal to (uv).

If $u'+2v'=u+2v$ (mod 5) this probability is one-half, for instance and if $u'+2v'\neq u+2v$ (mod 5) or $u'+3v'\neq u+3v$ (mod 5), the probability is zero. Complete specification of these probabilities would be tedious, and is not necessary for our purposes. From the strict point of view of randomization tests, it is clear that one can make only a test of the hypothesis of equality of treatment effects with a size of 50 per cent. In other words, one will use a test which will reject the hypothesis when it is true 50 per cent of the times. Surely in this case randomization cannot be appealed to as a justification for a normal law analysis.

ANALYSIS OF DESIGNS WITH NON-ADDITIVITY

It appears from the preceding description that by and large, experimental inferences may be based entirely on randomization providing additivity holds. The main concern of the experimenter should then be towards the determination and use of a scale for which treatments are additive in their effects. It has seemed to me that there has been in the past far too much emphasis on homogeneity of errors. It is remarkable that there is only one paper in the literature which tackles the problem of testing for additivity, namely Tukey's test [19]. This test is based on a linear model with normally distributed errors. Actually tests for homogeneity of variance tend to be used somewhat as tests for additivity, in that the error law which is guessed at is the basis for a transformation and an additive linear model is used on the transformed variable.

In view of the fact that perfect additivity on any one simple scale is unlikely to hold, it behooves us to consider the problem of estimation when there is non-additivity. The problem of the sensitivity of the randomization test, which will be discussed later for the case of additivity, is extremely difficult when there is non-additivity, and I know of no work at all on the matter.

For the case of randomized blocks it is found that block treatment interactions must be zero in order that the design be unbiased in Yates's sense.

It is perhaps worthwhile interpolating a consequence of this fact with respect to the design of randomized block experiments It does not appear to be at all desirable to section the experimental material into ordinary randomized blocks, of, let us say, highly different fertilities (or basal yields) because this procedure is likely to lead to block treatment interactions. In fact the requirement that the whole of the experimental material be as homogeneous as possible, which is the requirement in physical or chemical experiments, also holds for other experi-

ments. The impression is sometimes given that experimenters need only randomise and their troubles are ended. In fact only some of their troubles are ended.

It is easily proved that any comparison of observed treatment means is an unbiased estimate of the same contrast of the "true" treatment means. Here the "true" treatment mean is the mean which would have been observed if all the experimental material were subjected to the treatment. This seems a reasonable definition to use. Hence the inferences described here refer to the population of experimental units used in the experiment and not to some vaguely defined broader population.

Some work on the estimation of variances of treatment comparisons, based on a suggestion of Neyman [15], is given in [14], Section 8.3. If the non-additivity is slight in some sense, my own opinion is that the usual formulas are reasonable and unlikely to lead one astray. The situation with respect to non-additivity for the completely randomized design is essentially the same as for randomized blocks.

In the case of the Latin Square, however, there appears to be no work on non-additivity. The validity of the Latin Square for estimation is based entirely on the assumption of additivity, since one is picking out terms like $(x_{uv}-x_{u.}-x_{.v}+x_{..})$ at random to be attached to treatment totals. This is the fact which results in the difficulty of estimating both direct and residual effects in crossover designs

The randomization test for the Latin Square or for any randomized design is entirely valid in the sense of controlling Type I errors, but the approximation to this test by the F-distribution when there is non-additivity is apparently completely unknown.

THE POWER OF THE RANDOMIZATION TEST

It seems intuitively reasonable that the power of the randomization analysis of variance test when additivity holds, will be equal asymptotically (as r and t get large) to the power of the normal theory analysis of variance test. It seemed worthwhile ([14], Section 12.6) therefore to conduct a small empirical investigation of the power of the randomization test and to compare this power with the power of the analysis of variance test which is obtainable from Tang's tables [18]. There was an indication that the sensitivity with intermediate treatment differences is not well approximated by Tang's tables.

More work needs to be done on this matter.

CONCLUDING REMARKS

I have attempted to give an ordered and, in terms of present know-

ledge, an integrated account of the randomization theory of experimental inference. The gaps in this theory are in my opinion:

(1) the accuracy of the approximation to randomization tests by F tests
(2) the rather stringent role of additivity (which is also present in the case of normal law inferences)
(3) the power of the randomization analysis of variance test
(4) the consideration of alternative test criteria.

My own view on item (4) is that in view of the optimum properties of the analysis of variance test for the normal law model, the answer obtained in an investigation of this item will tend to support the use of the analysis of variance criterion.

There is, I suppose, no limit to the number of test criteria which could be used and in recent years there has been some emphasis on non-parametric criteria. It should be realized that the analysis of variance test with the F distribution has a fair basis apart from normal law theory and is probably in most cases a good approximation to the randomization analysis of variance test, which is a non-parametric test. When one considers the whole problem of experimental inference, that is of tests of significance, estimation of treatment differences and estimation of the errors of estimated differences, there seems little point in the present state of knowledge in using method of inference other than randomization analysis

REFERENCES

[1] Anscombe, F. J, "The validity of comparative experiments," *Journal of the Royal Statistical Society*, A-61 (1948), 181–211.
[2] Barnard, G A., "Discussion to paper by K D Tocher," *Journal of the Royal Statistical Society*, B-14 (1952), 91–2
[3] Cornfield, Jerome, "On samples from finite populations," *Journal of the American Statistical Association*, 39 (1944), 236–9.
[4] Daniels, H. G., "The estimation of components of variance," Supplement to the *Journal of the Royal Statistical Society*, 6 (1939), 186–97.
[5] Eden, T., and Yates, F., "On the validity of Fisher's z test when applied to an actual example of non-normal data," *Journal of Agricultural Science*, 23 (1933) 8–17.
[6] Fisher, R. A, "The goodness of fit of regression formulae, and the distribution of regression coefficients," *Journal of the Royal Statistical Society*, 85 (1922), 597–612.
[7] Fisher, R. A., *Statistical Methods For Research Workers* First Edition. Edinburgh: Oliver and Boyd. 1925.
[8] Fisher, R. A., "The arrangement of field experiments," *Journal of Ministry of Agriculture*, 33 (1926), 503–13

[9] Fisher, R. A., *The Design Of Experiments*, Edinburgh: Oliver and Boyd, 1935.

[10] Fisher, R. A., "Discussion of paper by Neyman, J. et al.," Supplement to the *Journal of the Royal Statistical Society*, 2 (1935), 107–54.

[11] Fisher, R. A., "The coefficient of racial likeness and the future of craniometry," *Journal of Royal Anthropological Institute*, 66 (1936), 57–63.

[12] Fisher, R. A., and Yates, F, *Statistical Tables*, Edinburgh: Oliver and Boyd, 1938.

[13] Grundy, P. M. and Healy, M. J. R, "Restricted randomization and quasi-Latin Squares," *Journal of the Royal Statistical Society*, B-12 (1950), 286–91

[14] Kempthorne, Oscar, *The Design And Analysis Of Experiments*. New York: John Wiley and Sons, 1952

[15] Neyman, J, with the cooperation of K Iwaskiewicz and St Kolodoziejczyk, "Statistical problems in agricultural experimentation," Supplement to the *Journal of the Royal Statistical Society*, 2 (1935), 107–54.

[16] Nissen, Øivind, "The use of systematic 5×5 Latin squares," *Biometrics*, 7 (1951), 167–70.

[17] Pitman, E. J G, "Significance tests which can be applied to samples from any populations III. The analysis of variance test," *Biometrika*, 29 (1937), 322–35

[18] Tang, P. C, "The power function of the analysis of variance tests with tables and illustrations of their use," *Statistical Research Memoirs*, 2 (1938), 126–49.

[19] Tukey, John W., "One degree of freedom for non-additivity," *Biometrics*, 5 (1949), 232–42

[20] Tukey, John W, "Dyadic anova, an analysis of variance for vectors," *Human Biology*, 21 (1949), 65–110

[21] Welch, B L, "On the *z* test in randomized blocks and Latin Squares," *Biometrika*, 29 (1937), 21–52.

[22] Yates, F, "The analysis of replicated experiments when the field results are incomplete," *Empire Journal of Experimental Agriculture*, 1 (1933), 129–42.

[23] Yates, F., "The principles of orthogonality and confounding in replicated experiments," *Journal of Agricultural Science*, 23 (1933), 108–45.

[24] Yates, F, "Complex experiments," Supplement to the *Journal of the Royal Statistical Society*, 2 (1935), 181–247

[25] Yates, F., "Incomplete Latin squares," *Journal of Agricultural Science*, 26 (1936), 301–15

STATISTICAL ABSTRACTS

All communications concerning this section should be addressed to the Abstracts Editor, Professor W. L. Smith, Department of Statistics, University of North Carolina, Chapel Hill, North Carolina.

Bross, I., "A confidence interval for a percentage increase," *Biometrics*, 10 (1954), 245–50.

Let there be two binomial populations with parameters p_1 and p_2. Then the relative excess (or increase) of p_2 over p_1 may be defined as $\Theta=100(p_2-p_1/p_1)$. Given samples of size n_1 and n_2 respectively, with number of occurrences x_1 and x_2, the problem is to find a confidence interval for Θ. The solution is obtained under the condition that p_1 and p_2 are sufficiently small, and n_1 and n_2 are sufficiently large to take x_1 and x_2 as having Poisson distributions. Then the joint distribution of x_1, x_2 may be written as the product of the distribution (Poisson) of x_1+x_2 with the conditional distribution (binomial) of x_1 given x_1+x_2. The binomial parameter in the latter distribution is $P=n_1p_1/n_1p_1+n_2p_2$; which is a monotonic function of Θ. The result quickly follows that $1-\alpha=P\{L<\Theta<U\}$ where L and U are defined by $L=100(n_1-(n_1+n_2)(U.L.))/(n_2(U.L.))$ and $U=(n_1-(n_1+n_2)\ (L.L.))/(n_2(L.L.))$ where $U.L.$ and $L.L.$ are upper and lower end points of a $1-\alpha$ confidence interval for P as defined above. Lincoln Moses, *Stanford University*.

Bross, I., "Misclassification in 2×2 tables," *Biometrics*, 10 (1954) 478–86.

In sorting the elements of a sample into two classes there are two possible kinds of error. Let the probabilities of these two kinds of error be called Θ and ϕ; they are "parameters of the classification scheme." If both are zero, then every individual is certain to be classified correctly. If not both are zero then in general the estimate of the binomial parameter p corresponding to the classification will be biased, and so will its estimated variance. When the same classification scheme is applied to samples from two populations and the equality of p_1 and p_2 is being tested the test remains valid (i.e., has the nomial level of significance) but its power is reduced. The estimate of the difference p_1-p_2 will be biased. Lincoln Moses, *Stanford University*.

Chernoff, Herman, "Rational selection of decision functions," *Econometrica*, 22 (1954), 422–43.

The author states as his fundamental purpose in this paper "... to see whether the theory of decision functions shows promise of being applicable to 'real' problems and not necessarily to specify how the theory is to be applied."

Following a few comments on "min max risk," "min max regret," and the "subjective approach" as alternative criteria for selecting a decision function, the author presents a set of ten postulates which are regarded as descriptive of a rational approach to selecting a strategy. The author points out that in his conception of the problem "... a rational criterion can be precisely described only in terms of the postulates. On the other hand, an intuitive notion leading to the selection of these postulates is the following. In a given problem, the statistician should first eliminate those strategies which are obviously bad. He should then dispose of some of the remaining which, while not so obviously bad, still fail to make the grade. After a certain amount of eliminating, the remaining strategies will be considered adequate. The statistician will have no reason to prefer any of these strategies to the others. The set of these strategies will be called the solution of the problem. It is *not* implied that the statistician necessarily considers that any two elements of the solution are equivalent." In the development of the set of postulates controversial postulates are indicated and those playing a critical role are discussed in some detail.

Postulates 1 through 8 are applied in the proofs of three theorems in a section labeled "main results." (This reviewer cannot claim to have checked the proofs.) Theorem 1 states that if a rational solution for the simplified formulation exists, randomized strategies are unnecessary. Through theorem 2 it is shown that regret matrices are relevant, i.e., the statistician may base his choice of strategy on the regret matrix. Theorem 3 (regarded as the main result) shows that for the class of all mixed problems involving n states of nature and a finite number of pure strategies, the unique

criterion satisfying the postulates formulated is equivalent to assuming that each state of nature has an a priori probability of $1/n$. Following the proofs of the theorems is a section given over to an interpretation of the results and a consideration of their implications with respect to the possibility of a rational approach to real problems via the decision function formulation.

Additional "miscellaneous" results are presented in a brief concluding section. IVAN M. LEE, *University of California*.

Cochran, W. G., "Some methods for strengthening the common χ^2 Tests," *Biometrics*, 10 (1954) 417–51.

The author cites two difficulties which commonly beset the use of the χ^2 test First, the test is often used where fairly clear alternatives to the null hypothesis are contemplated, but it is not tailored to take care of them; second, when the null hypothesis is rejected the question often remains, in what way does it fail to be true? The bulk of the paper is devoted to mitigating both of these difficulties. For testing goodness of fit to Poisson or binomial distribution the variance test $(\Sigma(x_i-\bar{x})^2/\bar{x}$ in the case of the Poisson) is recommended as ordinarily more sensitive than the general goodness of fit test; and when the latter is used the "no expectation less than 5" rule is probably unnecessarily strict, and wasteful of power. Analogously a test of skewness and/or kurtosis is ordinarily a more sensitive test of normality than the general goodness of fit test For each of these distributions methods are offered for testing that the observed frequencies are subject to linear regression (e.g., in time) or that there is a division into the first k "high" frequencies and the remaining $n-k$ "low" frequencies; and for each distribution there is given a method for constructing a single degree of freedom corresponding to any linear combination of the frequencies Contingency tables can be broken down into subtables for analysis. Methods and principles involved are discussed. In particular whether one should use an additive or nonadditive decomposition depends upon whether the table as a whole exhibits significance. Where one of the classifications has a natural order (such as degree of improvement) special procedures such as the use of "scores," are available, and natural. Methods of combining several 2×2 contingency tables are taken up. Where the data are affected with variation in addition to binomial or Poisson variation, then χ^2 analysis may be inappropriate; suitable transformation and application of the F test is likely to be preferable. LINCOLN MOSES, *Stanford University*.

Collier, R. O., Jr., "The least-squares analysis of a $p\times q\times r$ factorial design with unequal subclass frequencies," *Journal of Experimental Education*, 22 (1954), 297–83.

Using fixed main and interaction effects, Collier writes an interaction-zero model and a model "which attributes to each ijkth subclass a fixed effect," providing a "'pooled interaction' sum of squares " He then outlines the analysis of variance procedure and its application to a typing-reading experiment. JULIAN C. STANLEY, *University of Wisconsin*.

Durand, D., "Bank stocks and analysis of covariance," *Econometrica*, 23 (1955), 30–45.

In this paper the author presents results of one phase of a broader study analyzing bank stock prices by regression methods. The particular questions to which the author gives attention in the present paper bear on the validity of certain specifications of classical regression in the analysis of certain types of pooled cross-sectional and time series data. The type of data employed permitted suggestive tests of the specifications questioned.

The regression equation to be fitted in an analysis of the generation of bank stock prices is the following: Log $P=k'+b$ log B $+d$ log D$+e$ log E; where, $P=$ bank stock price, $B=$ book value per share, $D=$ dividends per share, and $E=$ earnings per share. The main phase of the analysis is based on annual data for 117 banks covering the period 1945–1952. One question considered is whether the total of 936 observations might be pooled for analysis to reduce the standard errors of the estimated coefficients. Pooling implies, of course, the assumption of uniformity of regression relations for different banks and for different years. The validity of this assumption is tested, employing methods of covariance analysis. For this purpose the 117 banks are divided into six groups, primarily on a geographical basis. A regression equation of the above form fitted to each year within each bank group provides 48 sets of regression coefficients in a 6×8 layout. This permits tests of the uniformity of regression relations between groups of banks within years and between years within bank groups. Also, on the basis of this design, residual variances were tested for homogeneity over years and over bank groups, using Bartlett's test statistic.

The author concludes that the above tests suggest heterogeneity in regression slopes and variances among groups of banks and, perhaps less conclusively, among years within groups. This in turn suggests that, in pooling to increase the number of observations, the stocks included within a group would have to be carefully selected for homogeneity and the time period covered would have to be fairly stable and relatively short. This led the author to explore the use of quarterly data as a possibility for increasing the effective number of observations yet restricting the time dimension to a relatively short period For this phase of the analysis, quarterly data for 15 of the 17 banks in one geographical group (New York) were employed. The data covered eight consecutive quarters beginning with the last quarter of 1951, thus giving a two-way 8×15 "experimental layout." The variance analysis model employed assumed uniform coefficients b, d, and e and assumed the constant term k' composed of three independent components: $k_{ij}'=k^* +a_{i0}+a_{0j}$, where k^* is unvarying, $a_{i0}(i=1,2,\cdots,p)$ is the influence of the ith bank, $a_{0j}(j=1,2,\cdots,q)$ is the influence of the jth quarter, and $\Sigma_i a_{i0}=\Sigma_j a_{0j}=0$. The resulting test suggests significant individual bank effects. Finally, the hypothesis of zero auto-correlation of residuals in different quarters is subjected to test. Two testing procedures are employed. One test was based on von Neumann ratios calculated for each bank from quarterly residuals derived from the deviations from regression in the cells of the 8×15 experimental layout referred to above. A second testing procedure (which is only mentioned) involved replicating price data within each quarter and testing for bank-quarter interaction. These tests lead to the rejection of the hypothesis of zero auto-correlation of quarterly residuals.

In a concluding section the implications of the test results are discussed briefly IVAN M. LEE, *University of California.*

Edwards, Daisy S., and Parkin, S. J., "Empirical investigation of the problem of disproportionate frequencies in analysis of covariance as applied to a methods experiment," *Journal of Experimental Education*, 22 (1954), 275–64.

For proportionate numbers of subjects in 18 method-school groups, F's were slightly smaller than when moderately disproportionate subclass frequencies, judged by the authors to be typical of educational experimentation, were used. Addition of a seventh school in which frequencies were quite disproportionate increased the "school" and "method x school" F's further. "Even with extremely disparate numbers . . . we obtain values of F for which the probabilities are very similar" to those for proportionate subgroups. JULIAN C. STANLEY, *University of Wisconsin.*

Epstein, Benjamin and Milton Sobel. "Sequential life tests in the exponential case" *Annals of Mathematical Statistics*, 26 (1955), 82–93

The authors describe sequential life test procedures when the underlying distribution is given by the exponential distribution. They assume n random items are available for life testing from the given distribution and wish to test $H_0:\Theta=\Theta_0$ against $H_1:\Theta=\Theta_1$ with Type I error $=\alpha$ and Type II error $=\beta$. The decision to continue the experiment is made as long as the inequality $B<(\Theta_0/\Theta_1)^r \exp -(\Theta_1^{-1}-\Theta_0^{-1})V(t)<A$ holds The constants, B and A, depend on α and β. $V(t)$ is a function of t which represents the total life observed up to time t. At the time the experiment is stopped, H_0 or H_1 is accepted according to violation of the first or second inequality. The authors, then, give two parametric equations which determine approximately the O C. curve and tabulate the approximate values of the expected number of observations required to reach a decision when Θ is the true parameter [viz., $E_\theta(r)$] for sequential tests of various values of $k=\Theta_0/\Theta_1$ and α, β

The authors also calculated exactly $L(\Theta)$ and $E_\theta(r)$ in special cases and obtained an upper and lower bound for them. They gave eight different problems, with solutions, to which these methods have been applied. A. E SARHAN, *University of North Carolina.*

Federer, W. T. and Schlottfeldt, C. S., "The use of co-variance to control gradients in experiments," *Biometrics*, 10 (1954), 282–90.

Using an illustrative example as a vehicle, an analysis of variance with quadratic covariance is given The method is presented as a possible alternative to the latin square design, in some connections. LINCOLN MOSES, *Stanford University.*

Graybill, Franklin, "Variance heterogeneity in a randomized block design," *Biometrics*, 10 (1954), 516–20.

In a randomized block design for comparing treatments two assumptions used in the analysis of variance may fail to hold in certain instances; there may be different vari-

ances associated with the different treatments, and the errors within any block may be correlated. The customary F test is then inapplicable, but Hotellings T^2 may be validly applied. A worked example is given. LINCOLN MOSES, *Stanford University*

Grundy, P. M., Rees, D. H., and Healy, M. J. R., "Decision between two alternatives—how many experiments?", *Biometrics*, 10 (1954), 317–23

The following problem is posed a new process will result in a gain of $(\eta-c)k'$ if put into practice. In the above expression k' is a known constant determined by the proposed scale of application of the new process; $(\eta-c)$ is the difference between increase in yield per unit and cost per unit The experimenter pays k units per experiment, he performs an initial experiment, obtaining an observation y_1 (assumed normal, with known variance σ^2) The authors, using the theory of the fiducial inference, show how to select n, the number of further experiments to be performed, in order optimally to balance the additional cost of experimentation against the expected gain due to taking the correct decision (i.e., adopting or rejecting the new process) LINCOLN MOSES, *Stanford University*.

Hamburger, William, "The relation of consumption to wealth and the wage rate," *Econometrica*, 23 (1955), 1–17.

The author is concerned with the relation $C=C(R)$, where C is per capita current consumption and R is per capita "lifetime consuming power anticipations" It is pointed out that the most commonly used index of R is current disposable income. In the author's formulation, a single measure such as disposable income is considered too comprehensive, i.e , it carries the implicit assumption that the relation of different types of income to anticipated lifetime consuming power are stable and identical In the formulation of the present paper, "human wealth" and "property wealth" are explicitly recognized as separate bases of lifetime consuming power anticipations. It is noted that measures of wealth rather than income are regarded conceptually as the more appropriate measures for reflecting consumer power anticipations.

A measure proportional to annual wage rate (kL), where L is annual wage rate, is used as an index of human wealth Property wealth is measured as the sum of two components $(G+jE)$ where G consists of net government debt to private sector, net foreign debt to private sector, and stocks of consumer durables, and E, an index of all other private wealth, is the sum of net interest, dividends, proprietors' income, and rental income. Specific sources and various adjustments introduced in constructing these measures are set out in an appendix.

A linear function was fitted by least-squares to annual data for the period 1929–41 and 1947–50. Written without explicit recognition of different components of consuming power anticipations, the linear relation is $C=aR+b$. Substituting for R the separate components indicated above this relation becomes $C=aG+ajE+akL+b$ Before fitting, this function was deflated by wage rate (L) giving $C/L=aG/L+ajE/L+ak+b/L$ The fitted equation provides estimates of the separate coefficients a, j, k, and b Estimates of a, j, and k were also obtained from a relation omitting b/L from the above relation (i e , on the assumption $b=0$)

Results are summarized and discussed briefly, directing attention to the suggestive character of the results from the substantive point of view IVAN M. LEE, *University of California*

Hayman, B. I., "The analysis of variance of diallel tables," *Biometrics*, 10 (1954), 235–44.

A powerful method of investigating genetical properties of several, say n, inbred lines is to construct all $n\times n$ single crosses (each line appearing once as male and once as female) and selfs, the resulting array of n^2 observations is called a diallel table. There is developed an analysis of variance model which permits testing for the existence of genetic variability among lines, dominance effects and effects associated with a line entering as male rather than female parent Methods of conducting large diallel crosses in latin squares and related designs are discussed. A clearly worked out 8×8 diallel cross is presented as an arithmetical example. LINCOLN MOSES, *Stanford University*

Kimball, A. W., "Short cut formulas for the exact partition of χ^2 in contingency tables," *Biometrics*, 10 (1954) 452–8

Convenient computing methods appropriate to decomposition of contingency tables are presented and illustrated An $r\times s$ table can be broken down into $(r-1)(s-1)$ individual tables, each of size 2×2. LINCOLN MOSES, *Stanford University*.

Li, C. C. and Sachs, Louis, "The derivation

of joint distribution and correlation between relatives by the use of stochastic matrices," *Biometrics*, 10 (1954), 347-60.

Two classical problems in genetics are to determine the probabilities for various genotypic combinations between relatives (of specified relationship) and to find genotypic correlations between relatives. The present paper, using matrices of conditional transition probabilities, gives a straightforward method of solving both problems, which are ordinarily quite cumbersome to deal with. LINCOLN MOSES, *Stanford University*.

Mandel, J., "Chain block designs with two-way elimination of heterogeneity," *Biometrics*, 10 (1954), 251–72.

A new class of designs, generalised from the recent chain block design, is presented. The assumptions are the ones ordinary to incomplete block designs. It is necessary only that the number of treatments be a multiple of the number of blocks. The designs are quite flexible and easily analysed. Both the theory and computations are clearly set forth in the paper. LINCOLN MOSES, *Stanford University*.

Moore, P. G., "A note on truncated poisson distributions," *Biometrics*, 10 (1954), 402-6.

An earlier paper considered a rapid less-than-efficient estimate of the parameter λ in which frequencies greater than or equal to s were not observed. The estimate there was $\hat{\lambda} = \sum_{r=0}^{s-1} rn_r / \sum_{r=0}^{s-2} n_r$ which is suggested by the identity $\sum_{r=0}^{s-1} r(\lambda^r e^{-\lambda})/r! = \lambda \sum_{r=0}^{s-2} (\lambda^r e^{-\lambda})/r!$ Three other types of truncation are considered in this paper and estimates of analogous type are offered. There is given some indication of the loss of efficiency associated with those (very easily calculated) estimates. The three additional types of truncation are:

a) Frequencies less than or equal to k not observed.
b) Only frequencies between s and k ($s<k$) observed.
c) All frequencies *except* those between s and k observed.

LINCOLN MOSES, *Stanford University*.

Nelder, J. A., "A note on missing plot values," *Biometrics*, 10 (1954), 400–1.

Where only one observation is missing in a balanced design, a convenient method of analysis calls for the construction of fictitious "observation" after which the complete-design analysis yields an unbiased error sum of squares. This invented value is only a computational device and "is not intended as estimate of the missing datum." The author points out that it, none the less, is an estimate of the missing datum, an unbiased estimate if the analysis of variance model is satisfied. Hence, if the invented value is absurd it stands as a warning that the model may *not* hold for the scale of measurement used, and that a transformation may profitably be sought for. LINCOLN MOSES, *Stanford University*

Rao, C. R., "Estimation of relative potency from multiple response data," *Biometrics*, 10 (1954), 208–20.

It may be in a biological assay that two (or more) graded responses are observed for each animal. The relative potency can be estimated from either series of responses. If they give essentially equal estimates then an optimum linear combination of the response variables may be sought in terms of which the estimation will be most efficient. This problem is treated for two response variables from a non-theoretical—but instead, fully computational—standpoint. This is done in such a way as to enable immediate generalization of computational methods to k-response ($k>2$) case. LINCOLN MOSES, *Stanford University*.

Sarkar, D. and Laha, R. G., "A modification of the variate-difference method," *Econometrica*, 23 (1955), 67–72.

In the classical variate difference method, a time series is assumed to be the sum of two components—a polynomial trend and a random element distributed independently with zero mean and finite variance. The modification in this paper is proposed to deal with time series in which prominent cyclical variations are present. It is assumed that the time series observations are a sum of a polynomial trend, a cycle of known period, and an irregularity, giving (1) $y_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_p t^p + A \sin((2\pi t)/\lambda) \; (+\alpha)) + \epsilon_t$; $(t = 1, 2, \cdots, N)$ where p is the unknown degree of the polynomial trend, λ is the known period of the cycle, and ϵ_t is a random component independently distributed with mean zero and variance σ^2. An unbiased estimate of σ^2 is derived. An estimate of σ^2 is also derived for the more general case in which the cyclical component is assumed to be the resultant of several harmonics of different but known periods $\lambda_1, \lambda_2, \cdots, \lambda_k$. It is noted that the generalised method can be applied if the λ_i are not known if it is known or can be assumed that $\lambda_i = \lambda/i$, $(i = 1, 2, \cdots, k \leq \lambda)$. Results of an application of

formulation (1) above for $\lambda = 7$ and $\lambda = 9$ are reported and compared with results of the classical method. The results suggest that the estimates of variance tend to stabilise at an earlier stage in the modified method. They also suggest that variance estimates are not highly sensitive to small differences in λ. IVAN M. LEE, *University of California*.

deVergottini, Mario, "Sugli indici di depressione," *Studi Economici*, n° 9/6, November, 1954, 445–50

To discuss the intensity of underdevelopment, one may use different indices. The usual index $I_1 = (a_1 - M)/M$, with a_1 per capita income of the depressed area, and M national income per capita, does not fit well to the problem. It is linked directly to the intensity of underdevelopment, but inversely to the number of people suffering it. In one country where the depressed area is very vast and M near the bottom, this index stays low and does not express the gravity of the problem It also does not allow comparison between two countries, because of its construction. Some simple indexes expressing the deepness of the depression $a_2 - a$, in reference to M, a, or a_2 are also used.

The index $I_3 = (a_1 - a_2)p_1/M$, with a_2 income per capita of the non-depressed area and p_1 the percentage of the total population in the depressed area, would be better. It is directly linked to the intensity of the depression, but also to the number of people in the depressed area. I_3 corresponds to the percentage of increase of the total national income that would allow to depressed areas per capita income in equal to per capita income in high-level areas. International comparisons are therefore not possible, I_3 depending upon the actual situation in each country.

Some other indexes can be constructed to take into account the fact that depressed areas are not monolithic. The index $L_4 = (a_2 - a_1)\ p_1p_2/M$ is of this type, giving the dispersion rapported to the mean.

Conclusion is that no judgment on depression can be made without careful examination of the indexes used for demonstration, the nature of index having a big impact on the results obtained. J. DAVID, *Conseil de l'Europe.*

Worcester, J. "How many organisms?", *Biometrics*, 10 (1954) 221–8.

In problems of biological assay involving quantal response, it sometimes happens that the dose is an aliquot from a suspension of organisms. The dose is the actual number of organisms applied, and is discrete, and subject to error, which may or may not be negligible as compared with the dilution-interval. Two general sorts of models in use for the problem are considered, and a third is introduced and illustrated. The standard logit and probit methods regard the dose as *not* subject to error, the "most probable number" method is based on the assumption that any animal actually receiving one or more organisms must then exhibit the response The model introduced assumes that the probability of response to an actual dose of x organisms is of the form $x/x+b$; the "most probable number" method corresponds to the use $b=0$. LINCOLN MOSES, *Stanford University.*

BOOK REVIEWS

Statistical Problems of the Kinsey Report. *William G Cochran, Frederick Mosteller, John W. Tukey* Washington, D C.· The American Statistical Association. 1954. Pp. x, 338 $5 00; $3.00 to members of the American Statistical Association. See review article on pages 811–29.

Colonial Social Accounting. *Phyllis Deane.* Cambridge, England: The National Institute of Economic and Social Research, Economic and Social Studies XI, 1953 Pp xv, 360. $10 00 See review article by William O. Jones, on pages 665–76

The Design and Analysis of Experiments. *Oscar Kempthorne* New York: John Wiley and Sons, Inc , 1952 Pp xix, 631 $8 50

GEORGE E P BOX, *Imperial Chemical Industries, Ltd.*

THIS is an extremely lucid and full account of the design of agricultural experiments and related statistical theory

The book contains 29 chapters The early chapters provide a discussion of scientific method and the principles of experimental design exemplified by Sir Ronald Fisher's famous tea-tasting experiment. There follows a brief but useful summary of elementary statistical theory, the normal distribution, and derived sampling distributions, and a discussion of such important concepts as orthogonality, estimation, and likelihood ratio tests A discussion of least squares theory and the general linear hypothesis appears next. Here the general results and proofs are set out in the compact and readily appreciated forms possible with matrix notation. A discussion of analysis of variance classifications then follows using the normal theory developed in previous chapters.

At this point randomization theory is introduced Following Fisher, Pitman, and Welch, the view is taken that the justification for the use of normal theory is that, for tests to compare means, normal theory supplies a good approximation to the result which would have been obtained had the full randomization test been performed. This view is sustained throughout most of the rest of the book which contains a full account of the standard agricultural designs: randomized blocks, latin squares, factorial experiments, split plots, fractional replication, lattice designs, etc.

The author refers to the randomization test model as the finite model, the normal theory model as the infinite model This is perhaps not a happy distinction for many, including the reviewer, would prefer to regard the randomization test as having an infinite reference set

For the reader who wishes to apply the techniques in fields other than agriculture, some caution is necessary. It has sometimes been rashly assumed that the strategy of experimentation specifically developed for agricultural application is directly transferable into all other fields. This is not so.

The basic characteristics of agricultural experimentation are· (1) A long time (about one year) must elapse between the planning of an experiment and the obtaining of the results, which then all become available together.

(2) The experimental error is large Since in these circumstances another year must elapse before questions still in doubt can be resolved, large 'comprehensive' designs are employed.

By contrast, the basic characteristics of much industrial experimentation are: (1) Experiments are usually performed one after the other and comparatively rapidly. (2) The experimental error is often somewhat smaller than in agriculture. In these circumstances sequential experimentation is the natural procedure. The results of a comparatively small group of trials are used to plan the next set and each subsequent set is planned in the light of all the knowledge available up to that time. In this way it is ensured that no more observations are made than is necessary to obtain the desired precision, and furthermore that trials are performed at levels of the factors which are of real interest to the experimenter. To ignore the possibility of a sequential strategy when such is possible may result in a great loss of efficiency and confirms the experimenter's worst fears concerning the good sense of the statistician.

This point is illustrated by the example on page 426 of the book This describes an experiment to determine the effect of 7 factors thought to influence the readings given by a machine designed to test the consistency of tinned foodstuffs A 1/9th replicate of a 3^7 factorial design was used, and this involved 243 observations After the experiment had been performed it was apparent that nothing like this number of observations was necessary either to attain the precision desired or to estimate complex effects A sequential strategy beginning, for example, with a high order fractional replicate of a 2 level factorial design would have provided a more efficient and natural approach and would have achieved the desired result with a small fraction of the effort. This point is emphasized because it is believed that outside the field of agriculture the sequential situation is by far the most common one.

These reservations should not obscure the fact that here is an excellent text book of value to student, teacher and practical worker alike.

An Outline of Biometry. *C I. Bliss and D W Calhoun* New Haven, Connecticut· Yale Co-Operative Corporation, 1954. Pp. x, 272, xvi, plus 18 tables and 5 figures. $4 50 plus 30¢ postage Paper.

H. FAIRFIELD SMITH, *North Carolina State College*

AT A Galton Laboratory tea in 1937, when there were few text books to guide a student in study of statistical methods for research, Fisher remarked that the way to obtain a good one would be for everyone who might feel the urge to try his hand and see which product would survive Even then the prospect of the volume of literature such a procedure might produce somewhat dismayed me. The flood is now upon us. Fair comparison is a task for a specialist in nothing but statistical texts. Some years ago I gave up in despair trying to index, let alone be acquainted with, current production; therefore I must disclaim ability to give accurate rating to any one. However,

An Outline of Biometry should take a high place in the race for survival.

The tone of the work is set in the preface which remarks: "By providing a working manual of lecture notes, essential formulae, numerical examples and statistical tables, we hope that the student can mimimise his note-taking and give greater attention to class discussion. The Outline reflects two tenacious principles of my teaching: that a course well stocked with tested alternatives is the most effective compromise between the breadth of the field and the time limits of the graduate program, and that the only way to grasp basic concepts is to calculate numerical examples with actual data". Excellent precepts and they are admirably executed! The book is intended for an elementary course for biologists extending to 40 lectures accompanied by four hours of laboratory to each lecture. Although described by the authors as a lecture manual its close gearing to practice makes it rather a laboratory recipe manual. Yet it does not read as stiltedly as that description might suggest owing to the logical development from one topic to another and the skill with which the recipes are collated. Bliss states that the unusual format was adopted because as a student he found a formal course outline to be an effective mode of study. Although one might anticipate such formulation to be deadly in a text book, the way this outline succinctly drives home each procedure looks like convincing evidence for the effectiveness of the method.

The work begins at a very elementary level with two pages of pointers in arithmetic which anyone reaching a statistics course should know from school, but statistics lends new eyes to a biologist's arithmetic and the refresher may be worth while. Anyone who needs to be told, however, the order of arithmetical operations will not two pages later know the meaning of a derivative stated without explanation. Such inconsistencies seem inevitable in elementary statistics texts, and presumably a reader of the first chapter is intended to skip the odd paragraphs where he is invited to do partial differentiation; the practical instructions are still clear without it. The subsequent computing instructions are both useful and needed.

The subject matter is, in the main, that usually covered in an elementary course: starting from the probability interpretation of the classic tea tasting experiment it proceeds through the binomial distribution, χ^2, normal distribution and tests based on it, interval estimation, analysis of variance, simple regression and correlation, to Poisson and negative binomial distributions. The authors regret the lack of extension to covariance, partial regression, discriminant functions, probit analysis, sampling and sequential analysis. But covariance and sampling are the only ones that need be regretted. The others are probably better omitted from the kind of course envisaged by this book, and left to a more mature level.

No attempt is made to teach theory, although the source whence methods derive is lightly indicated. At the same time, so differently from many elementary texts, it is clear that the underlying theory is thoroughly understood by the authors themselves. Difficulties are not slurred over. An out-

standing example is the statement in elementary language of the distinction between confidence and fiducial intervals, a statement which puts to shame many texts, superficially more erudite, which pass this by in a cloud. With absence of theory goes any statement of maximum likelihood which appears only as an adjective for some statistics and in connection with a brief plunge into amounts of information associated with transformed metameters. This section must leave gasping those for whom Chapter 1 was written, may stimulate some to delve deeper.

Methods can seldom have been presented more concisely; there are only 145 pages of text. With the help of a neat table of required F values, the Bross fiducial limits are stated in five lines. Orthogonal polynomials are presented in one and a half pages, and that includes a derivation for unequally spaced values! Inevitably such compression leads to some inconsistency, as in the instruction to solve the general polynomial form by using powers of x as independent variables, although multiple regression and its normal equations have not been given. Presumably it is merely an indication to a worker to look up the reference cited.

Description of models I and II for analysis of variance pleasingly maintains Eisenhart's basic distinction which lies in the kind of parameters to be evaluated: of location or of dispersion. It states simply that model I directs attention to the means of "particular treatments of interest to the experimenter," and avoids the later superimposed idea of "fixed effects" whose whole population is present, a concept which confuses the student who complains that he can think of myriad other treatments essentially belonging to the same population Exposition of the correlation coefficient is happily restrained to sampling from bivariate normal populations for which it is an appropriate statistic and thus avoids suggesting uses, all too common, for which it is unsuited. Two examples reasonably conform but example XV-6a is a regrettable exception Here the distribution of y is distinctly skew, and the regression of x on y is curved (although that of y on x may be linear), so that these data are not appropriately described by the correlation coefficient.

Transformations get a larger than usual share of attention Three and a half pages describe many by which curves may be converted to linear forms, including probits and logits; and a chapter of eleven pages is devoted to their uses for making numerous types of data amenable to analysis of variance. The utility of regression with transforms is developed with a chapter on bioassay when parallel linear regressions may be assumed. Numerous non-parametric and "quick and dirty" methods are described. Since arrangement is by types of data, these jostle with standard procedures instead of being relegated to another part of the book as if constituting a different subject. The logical grouping thus achieved may make the book useful for reference by workers who do not need it for information but who have not yet accustomed themselves to think of these newer methods in their appropriate places in their battery of tools. Some may consider tests given

for gaps in a group of means as too dogmatic, but in the present confused state of that subject they may be as good a choice as is practicable without more discussion than would conform to the style of this book. A further plunge is made in favor of certain "pool" rules which are not yet usually regarded as tried recipes. It might be advisable to indicate that bias is only locally minimized for some gain of power, the absolute minimum of course being at "never pool."

Convention dictates some criticism if only to prove that the reviewer has not read uncritically! In a book so brief and practical it seems redundant to describe association coefficients and corrections for bias in the standard deviation: so far as I know the former have never yet served any useful purpose and the latter are never used. However they occupy together only three-quarters of a page. The first page of Chapter XII rather confuses the relations of regression with structural and functional relations. The first sentence of paragraph Ala seems to describe the Berkson case, yet that is introduced at the end of the paragraph as something different. In Chapters X and XI formulas for sums of squares between groups in terms of means appear incorrect unless Page 86 has been remembered with exceptional care $S(\bar{y}_g-\bar{y})^2$ is intended to imply summation over all N values with repetitions as entered in, for example, Table d of example X-3; whereas in the formula to which it is set equal ST_g^2/f implies as usual summation only over the number of groups. In example XII-12 under M.S. for 284 45 read 248.45 (if not noticed the source of some of the F ratios may seem puzzling).

At page 196 delete $(N+1)$ from the denominator of formula XV-$D3b2'$ to test deviation of $\bar{x}$, $\bar{y}$ from μ_x, μ_y in a bivariate normal sample. For students who wish to look further and see whence formulas derive some indication should be given that this is the F transform of Hotelling's T^2. Formula 1′ for the corresponding test if μ_x were known (again with the erroneous $N+1$) seems to have been derived from 2′ by setting $\bar{x}=\mu_x$ Surely this is incorrect? Knowing μ_x does not mean that a bivariate sample has $\bar{x}$ equal to it. Maximum likelihood indicates that under the given conditions the best estimate of μ_y is $\bar{y}-b_{yx}(\bar{x}-\mu_x)$, which presumably would be tested with the usual t. It may have been intended to suggest that a sample was drawn such that $\bar{x}$ was deliberately made equal to the known μ_x, but this would return one to the ordinary regression case and a t test.

An interesting collection of 18 tables and 5 figures contains some which are not usually given in text books, such as 5 per cent points of $F_{max}=s_{max}^2/s_{min}^2$ in a set of k variances, range tests for cross classifications, and a chart for standard error of the mean in truncated samples from normal distributions. The nice chart of χ^2 originally published by Bliss in 1944 is included. A hundred and fourteen examples from real data must constitute one of the finest such sets of biological data ever assembled. The bibliography lists 41 well chosen texts and books of tables, and 240 miscellaneous citations, the majority of which are the sources for examples culled from an impressively wide field. This list of citations introduces a feature for which

I have often wished in bibliographies but never before seen put into practice: it indicates the page on which each citation is referred to. The book is produced by very clear multilith printing on one side only of $8\frac{1}{2}\times 11$ inch pages. This makes it bulky but leaves room for working notes. Cardboard covers of the wire-stitched review copy would seem unlikely to survive hard laboratory use; but the more popular form is punched for use in loose-leaf binders. We are informed that few copies are now left in stock. A revised and expanded edition is expected to be published in book form in about ...e years.

If told another elementary text is to be written, my reaction is: Please, not another! But *An Outline of Biometry* is a well worth while effort and should find a useful place in teaching statistical methods.

Statistics and Mathematics in Biology *Edited by O Kempthorne, T A. Bancroft, J. W. Gowen and J. L Lush* Ames, Iowa The Iowa State College Press, 1954 Pp ix, 632. $6.75

J. H. Bennett, *Cambridge, England*

This publication comprises forty-four papers delivered during a five-week conference on Biostatistics held at Iowa State College in 1952. The conference, which was organized by the editors of this volume, grew out of "the evident need for synthesizing the concepts and methods of biology with the concepts and methods of statistics and mathematics" and was the means for bringing together many biologists and statisticians, including some distinguished scientists, to discuss problems in quantitative biology.

The text is arranged in five parts corresponding with the different topics considered during the five weeks of the conference. I. General Biometrical Principles and Procedures; II. Changes in Population Number; III. Estimation of Populations; IV. Determination of Biological Response, and V. Genetical Analysis of Populations Some but not all of the biological and statistical chapters are paired. As is fitting, this work is opened by Snedecor with a brief outline of the historical background of biometry. He is followed, first by Wright, who presents a résumé of some of his early work on path analysis, and then by Tukey who reminds us that regression is generally a more useful concept than correlation. Then Hotelling tells us of some contributions, mostly his own, to multivariate analysis. Problems of classification on the basis of multiple measurements are taken up by Isaacson and others in the next three chapters and this first section concludes, after a paper on the fitting of growth curves, with four chapters dealing with general questions of experimental design These include two short review articles and a paper by Quenouille on the use of genetically homogeneous subgroups in experimentation.

Part II comprises three chapters describing experimental and mathematical studies of competition between species.

Part III, on the estimation of population number, is opened by Jessen with a useful survey of the methodology of sampling human populations. The other seven chapters of this section deal with the methodology and some practical difficulties arising in forest inventory and in estimating the size of fish, wildlife and insect populations.

Part IV begins with four chapters on bioassay; Cornfield's lucid account of the comparison of toxicities with quantal responses and Bliss' thorough discussion of insecticidal assays being two that are noteworthy. Six chapters on sensory tests for food products and nutritional and behavioral responses of animals make up the remainder of this section.

Part V contains eleven chapters on subjects related more or less to population genetics. After a brief review by Levine of the genetics and racial frequencies of some human blood groups, Cotterman gives an account of the estimation of gene frequencies in populations. In two chapters, Neel and Schull write about various aspects of the genetic studies that have been made on atomic bomb survivors. Pollard describes the physical approach to the study of the biological effects of radiation on the living cell and Gowen writes about his radiation experiments with viruses. Outstanding chapters are contributed by Griffing who carefully explains the statistical and genetical analysis involved in a problem of quantitative inheritance n tomatoes and by Crow who brings together various approaches to the genetical analysis of small populations and considers different ways of calculating an effective population number. The book ends with a chapter on the role of the nucleo-proteins in cell growth and division.

The length and treatment of these chapters differ as much as their subject matter and authors. A wealth of material is scattered through the book even if, in some places, its distribution is rather thin. Many of the papers will provide admirable discussion material for graduate students interested in quantitative biology.

The conference itself, although devoted, it would seem, to a field that was much too wide, must surely have given a direct boost to the use of quantitative methods in biology. But if the synthesis of the concepts and methods of biology with those of statistics and mathematics means anything more than bringing together all these papers, it has not been achieved in this book.

Not all papers delivered at conferences have lasting value or deserve preservation. If this were more widely realized, volumes such as the one under review would be much smaller and no less valuable.

Statistical Analysis in Chemistry and the Chemical Industry. *Carl A. Bennett and Norman L. Franklin.* New York: John Wiley and Sons, Inc., 1954. Pp. xvi, 724. $8.00.

F. R. Himsworth, *Imperial Chemical Industries, Ltd.*

It is well, before reviewing a book, to consider for whose benefit it is written; the present volume is addressed to "those in the chemical industry whose interest in this subject has quickened in recent years." It will be

found interesting and stimulating by those chemists who have made a more than average study of statistics, and by statisticians working in industry, but not by chemists who are mildly interested, or who want to learn elementary statistics; these will find the book very heavy going, and will be repelled, rather than helped, by the masses of algebra used in deriving sampling distributions, expectations of mean squares, and regression formulas. This is not a criticism of the book: it is intended to warn a class of reader who would not appreciate it, and who would, if he survived the discussion of probability in Chapter 3, give up in Chapter 4.

Chemists who are mathematically inclined, and who have already read widely, will find this book very helpful. They will find in it derivations of the tests they have hitherto taken on trust; they will find useful accounts of order statistics and non-parametric tests which will enable them to escape from the dilemma of what to do if the data are not normally distributed; and many techniques which will occasionally be useful, but which are otherwise available only in journals. They should, however, have a clear idea of the practical applications of statistics, and of the action to be taken after the statistical analysis has been completed, before they attempt to master the book. There is a strong tendency in the many examples to carry out a detailed statistical analysis, ending up with conclusions such as "all main effects are significant." This may illustrate the particular technique being discussed, but it would be much more helpful if these conclusions were translated into practical guides for action, even in a book of this nature.

A very sound piece of advice to anyone about to analyze results statistically is to look at the data, summarizing them in sub-tables and graphs; this will often tell him, at least qualitatively, all he wants to know. It may tell him that two factors interact strongly, in which case an analysis of the whole body of data is often valueless. The example on p. 395 illustrates this point admirably. Simple inspection conveys all the necessary information, and leads to conclusions more complete and more precise than those given. It may be said that the object of the example was to illustrate a method of analysis, but any example should give the best treatment of the data, and should be used to illustrate a technique only if this technique is the appropriate one. This criticism can be made of many of the examples. It is only when writing a textbook that one realizes how few suitable examples there are for illustrating a simple technique. The fact is that in practice, matters are rarely simple enough for the straightforward application of a standard technique, a warning which might well be given in texts in industrial statistics.

Chapter 2 deals with the assembly, grouping and graphing of data, and with simple statistics of location, dispersion and, perhaps rather prematurely, regression. It is stated on p. 20 that the sample standard deviation, mean deviation, and range are all biased estimates of σ. The first will be found puzzling without some explanation, and the bias is so slight that the statement is better omitted. To say that mean deviation and range are "biased estimates of σ" is playing with words, and distinctly unhelpful.

Here, and throughout, too many significant figures are retained in the final answers.

Chapter 3 on "Probability and Samples," although not very clear to a newcomer to the subject, gives the fundamental principles on which later arguments are based. The independence of variables is briefly discussed, but the chemical example is badly chosen; nitrogen content and nitrogen/phosphorus ratio are unlikely to be independent The matter is further obscured on p. 53, where the original weight of a coal sample and its weight after drying are said to be independent! Independence of errors is intended. The use of "ave (x)" = "average value of x" in place of the usual $E(x)$ = "expected value of x" has little to commend it, and leads to some curious expressions, such as "the average value of a single observation," "average mean square," etc. "Mean" is sometimes used instead of "average," and "expectation" in at least one place.

Chapter 4 on "Mathematical Machinery" will be interesting to the enthusiast; it deals with generating functions, semi-invariants and other matters to be found only in works on mathematical statistics, and gives full derivations of the important sampling distributions. The letter "t," which used to be an estimated mean divided by its estimated standard error, is sadly overworked. We have t, t_r, t', t'', t''', and maybe more.

Chapter 5 on "Statistical Inference" covers confidence limits and tests of significance along the usual, and some unusual, lines. It is unnecessarily befogged by the inelegant shorthand of probability theory; plain language would occupy little more space, and would make the situation clearer. The method given for finding the number of trials required, for example, is obscure, and should be completed by finding the acceptance values for the null hypothesis. A curious suggestion is that the z-test, assuming z normally distributed, is quicker than the F-test. The section on "Sequential Tests" is inadequate.

Chapter 6 deals with Relationships between Variables, i.e. regression and correlation, along conventional lines, with the usual complicated algebra and arithmetic. Regression always attracts the beginner, since apparently he can extract all the information from any old data. In fact it bristles with traps for the unwary; multiple and curvilinear regression should be approached only by experts, and with the utmost circumspection. Simple methods should always be used before an elaborate calculation is undertaken. A "significant" value of r may be quite unhelpful, such as that on p. 280, whose 95% limits are 0.05 and 0.92! The final sections on "Correlation of More than Three Variables" and "Discriminant Functions" are not likely to tempt the reader to use these techniques, which is perhaps as well; these are games for professionals (if anyone) to play.

Chapters 7 and 8 on "Analysis of Variance" and "Design of Experiments" occupy 280 pages—a fair sized book in itself—and are, in the reviewer's opinion, the most useful and interesting of all. They could well be expanded, subdivided, and made into a separate volume. In addition to the conven-

tional treatment, there is much detailed discussion on such matters as "variance components" and confidence limits for these, co-variance, significance levels when a number of mean squares are to be tested, and other thorny problems not usually mentioned in books on applications of statistics. Too much emphasis is placed on the testing of null hypotheses, though in some cases this is supplemented by the calculation of confidence limits, little or nothing being said about the power of the tests.

It has always seemed to the reviewer that for an experimenter to choose his conditions carefully, carry out an expensive experiment and a lengthy analysis, and then to apply a test which starts from the hypothesis that the experiment has been a complete failure, is somewhat illogical, and reflects great lack of faith in his scientific knowledge. It is usually inconceivable that the null hypothesis, $\sigma^2=0$, could be true; so why test it? What do we do if it is not contradicted? This book goes a certain length in the right direction, but the F-test is solemnly applied to test a hypothesis which cannot possibly be true, and which is obviously untrue from a rough inspection of the data. It is a good rule never to carry out the formal analysis without first looking at the data and drawing any obvious conclusions In many of the examples the conclusions finally reached are obvious, and in some cases the obvious conclusions are not reached at all. When a large interaction occurs in a crossed classification there may be no point in analyzing the data as a whole, two entirely separate sets of conclusions may be required, each applicable to one level of a factor. An excellent example is that on p. 395, on the testing of coal for volatile matter in steel and silica crucibles One coal gave fair agreement and the other hopeless disagreement, and the data as they stand are not worth analyzing. From the data on p. 368, all the conclusions can be arrived at very simply, without analysis of variance, which only obscures the issue. A check on the residual sum of squares, from the differences between duplicates, reveals an error, apparently in the table of results. The importance of checking calculations is not sufficiently emphasized; simply repeating the calculation is not good enough.

These chapters will repay (and will need) careful study. There is a real attempt to face the many difficulties which are usually ignored. They would be clearer if allowance for random samples from finite populations had not been made so regularly, leading to much very cumbersome algebra which, incidentally, contains several mistakes, and if interaction terms had not been included in the expectations of the mean squares, only to be underlined, the underline meaning "omit for a Type I set-up, which this is," as the reader will discover after he has been puzzled somewhat.

Chapter 9 on "Analysis of Counted Data" and Chapter 10 on "Control Charts" follow conventional lines and call for no special comment. Chapter 11 on "Some Tests for Randomness" may not be found particularly useful, and the example on p. 674, in which the significance test contradicts the common sense view, is not helpful, particularly when two other tests strongly support common sense.

A review tends to contain more words of criticism than of appreciation, sometimes rightly so. This book can, however, be strongly recommended to all serious students of applied statistics. Because it breaks so much ground that is new, so far as textbook exposition is concerned, it inevitably invites criticism on points of detail. It is a real attempt to introduce new and less well-known techniques to the industrial statistician. Some of these may not stand up to the requirements of practical usage, but they deserve to be better known and to be tried in practice. Non-parametric tests are a case in point. The reader will also find that after digesting the book he will understand the standard techniques better than before, and will appreciate some of their shortcomings. He may be in some respects a sadder, but he will certainly be a wiser man.

Printing and lay-out are good, though some tables seem to be unnecessarily separated from the accompanying text. The book has a fair number of errors and misprints. An annoying habit is to use "1," "2," etc. in the text rather than "one," "two," etc.

Probleme der Statistischen Methodenlehre in den Sozialwissenschaften. *Oskar Anderson.* Würzburg: Physica-Verlag, 1954. Pp. 345. 16.00D.M.

WERNER Z. HIRSCH, *Washington University*

This text offers a concise treatment in the German language of many of the concepts and methods of modern statistics and is designed to meet the needs of a second or intermediate course in statistics for social scientists, especially economists. The author expects those who use the book to have the mathematical preparation of the average German high school graduate and no fear of the algebraic formulations of ideas.

The book is organized into nine chapters: (1) Discussion of statistical method, (2) Descriptive measures, (3) Index numbers, (4) Statistical errors, (5) Probability theory, (6) Decision making, (7) Time series analysis, (8) Correlation, and (9) Logic of statistical method. Since the chapters vary greatly in length, the orientation of the book is perhaps best made clear by stating that the beginning one-third of its material is descriptive, while the rest deals mainly with problems of inference. This reviewer has no quarrel with the order and organization of the material. Not only does it have the advantage of putting easier things first, but also it is quite logical. Yet, many statisticians may not agree with the balance and inclusiveness of the presentation.

Following a brief discussion of the arithmetic mean and standard deviation, the author treats index numbers in a very effective manner and in great detail. This chapter proves especially instructive because, after a rather conventional treatment of the main problems and formulas of index numbers, there follows a stimulating discussion of some of the theoretical and practical problems of cost of living and production indexes.

The next section is proof of the author's experience and his ability to use his mathematics admirably well. His idea is to prove rigorously a few key concepts and to derive from them a number of important statistical rules and formulas. From a detailed discussion of probability flows the binomial theorem and the Poisson distribution, from which in turn are derived the rules and formulas for making confidence interval estimates and for testing hypotheses. The logical flow of this section is excellent. Many statisticians will be disappointed, however, that, only in introducing the topic of inference brief mention is made of errors of types I and II. The presentation of significance tests is concerned exclusively with making decisions of a known probability of committing an error of type I, and errors of type II are altogether neglected.

The chapter on time series analysis may perhaps be criticized more than any other chapter, from the point of view of balance. Following some general introductory remarks, the conventional least squares method of finding trends is presented. Since the author helped develop the variate-difference-method, he spends considerable time on this rather infrequently used technique In a few pages and without examples he disposes of the problem of seasonal variation, without indicating what to do about irregular variations and how to isolate cyclical variations. The applications to which time series analyses can be put are left very much in the dark. An omission, though of less importance, is the lack of significance tests for trend and seasonal variation. For a book which addresses itself to social scientists in general and economists in particular, a more appropriate section on time series analysis could have been written.

The final and important section of the book, dealing with correlation, is very well written With a minimum of mathematics the relations between correlation and regression analysis are shown, and the various formulas are derived and applied. The inferential aspects of correlation are ably presented. The only complaint that one might have is that of the sixty pages of the chapter on correlation, only ten are devoted to multiple correlation.

This reviewer believes that the inclusion of a discussion of decision making when there are more than two choices—analysis of variance—could have rounded out the book. Also, the concepts of estimation and decision making might have been briefly extended into the important areas of sampling methods, which are becoming so very important as tools of the social scientists.

Many statisticians are wondering what new contributions to the field are being made in other languages, especially German. In this connection it is interesting to note that in summing up the state of knowledge of statistics in general, and of statistics for social scientists in particular, the author draws almost exclusively on the work done outside Germany. His detailed index of authors comprises about one-and-one-half times as many Anglo-Saxon as German references.

In summary, this book is a valuable addition to statistics texts in the

German language. It is well written and the material in it is up to date. There can be no doubt that it will greatly contribute to making available to those who read only German much of modern statistics as it has been developed and taught abroad.

Probability Theory. *Michel Loève.* New York: D. Van Nostrand Company Inc., 1955. Pp. xv, 515 $12 00.

Walter L. Smith, *University of North Carolina*

It appears to be a growing practice for authors of books on statistics to provide in their prefaces a suggestion concerning the way in which readers should study their works. Recently, Leonard J Savage recommended us to read about the foundations of statistics sitting bolt upright on a hard chair, at a desk, and now Loève asks us to approach his monumental treatise on the foundations of probability theory "armed permanently with patience, pebble, and reed". This advice is to be taken seriously, for the 515 pages of this masterly treatise contain an enormous wealth of material much of which is presented, of spatial necessity, in a very concise style, and, moreover, Van Nostrand have found it desirable to resort to a smaller type than usual, which makes this book somewhat more difficult to read (in the physical sense) than their earlier books in this series.

Loève's book, then, is an authoritative account of probability theory as it exists at the present time, and the dust jacket claims accurately that, in the class of books dealing with probability theory, "there is nothing as comprehensive in treatment, . . . , or as up-to-date". But this is definitely a book for the specialists Many highly successful practicing statisticians will dispute Loève's claim that Part One of his book contains the notions of Measure Theory that every statistician requires. (Has Fisher ever used the Radon-Nikodym theorem?) Few statisticians with less than powerful probabilistic tendencies will feel prepared to pay the high cost of this book and devote to it the close study it most certainly deserves. Nevertheless, those who are primarily concerned with probability theory have every reason to be grateful for the present work, which is destined to become a standard reference in the subject, and which fills so adequately an unfortunate gap in the literature of probability (recently lamented by Doob in his "Stochastic Processes").

The first 50 pages are devoted to an introductory treatment of familiar "elementary" problems in probability theory, which do not require the use of measure theory. The purpose of this section is to introduce in a natural way the basic ideas involved in probability theory. It does not replace the sort of treatment one finds in, say, Feller's "An Introduction to Probability Theory", but forms a useful and compact revision course and introduction to the notation to be employed (it includes a discussion of Markov Chains on the lines of Kolmogorov's classical paper). A similar remark applies to Part One, which fills the next 94 pages and is devoted to Measure Theory. In spite of the existence of a few excellent and readily available

treatises on Measure Theory, it is perhaps desirable to have a collection of the most important measure-theoretic results included in a treatise on probability theory. Loève provides us with more than this, he supplies a "self-contained" course on measure theory. If the reader has already obtained a grounding in measure theory (such as may be obtained by reading Munroe's "Introduction to Measure and Integration") then Part One becomes, like the introductory section, an excellent revision course. But the level of conciseness that Loève has aimed at has entailed a somewhat sophisticated attitude which a reader meeting measure theory for the first time might easily find trying.

The thorough-going measure theoretic treatment of probability theory begins on page 149 and occupies the remaining 366 pages. It is divided into: Part Two, on the general concepts and tools of probability (74 pages); Part Three, on sums of independent random variables and their limit properties (114 pages), and Part Four, devoted to a careful explanation of conditioning, to limit properties of sums of dependent random variables, and in the last section to random functions of the second order. An enumeration of chapter headings would be tedious, but certain chapters deserve especial mention. Chapter V on sums of independent random variables is excellent. Chapter VI, on the problem of convergence of laws of sequences of sums of random variables deals not only with the classical theorems of the Central Limit type, but with the problems of more recent origin concerning the determination of the family of all possible limit laws and of the conditions for convergence to these laws. This chapter gives a welcome exposition of a subject which, until recently, one could only study in papers scattered through many journals (and mostly in languages other than English). For similar reasons, it is good to find that Loève has included an up-to-date account of Ergodic Theory as Chapter IX.

There are many brief historical sketches of the development of various problems and of their solutions, a feature which will appeal especially to newcomers to the subject. There is also a very adequate supply of examples ("complements and details") at the end of each chapter. Unfortunately, several misprints occur in this first edition, but the reviewer learns that an errata sheet is now being provided, and these errors will doubtless be eliminated in any later editions

Every serious probabilist should, and doubtless will, possess a copy of this important work. Loève is to be complimented on completing his Herculean task at a uniformly high level of elegance (a task which, one gathers from his preface, he regards as an essay in the poetic form!).

A Million Random Digits with 100,000 Normal Deviates. ***RAND Corporation.*** Glencoe, Illinois: The Free Press, 1955. Pp xxv, 400, 200. $10 00.

This is by far the largest and best collection of random digits yet (or likely to be) published in book form It is the collection from which the 22,475 random digits published in this Journal from 1952 to 1954 were taken.

The million random digits are arranged 2500 to a page on pages numbered 1 to 400. There are fifty numbered rows on a page, each row divided into blocks of 5 and the blocks set off in pairs. In addition there are 100,000 unit normal deviates of 4 significant figures (3 decimals) and a plus or minus sign. These have been derived from 500,000 of the random digits in a manner and sequence which is fully described. They are arranged 500 to a page, 10 in each of 50 numbered lines, on pages numbered 1 to 200.

The 25 page Introduction, which regrettably is unsigned, is a worthwhile piece of statistical literature itself. The method of producing the random numbers is described, and a series of interesting tests of their randomness is presented. The instructions for using the tables are excellent; they include simple and clear directions for generating from the random digits random numbers from any specified population.

The book is excellently printed and bound, and the price seems reasonable for the size. It is a book which most statisticians will want to own, and if the history of previous tables of random numbers is any criterion, it may not be in print long

W.A.W.

A History of the Faculty of Political Science, Columbia University. *R. Gordon Hoxie, Sally Falk Moore, Joseph Dorfman, Richard Hofstadter, Theodore W. Anderson, Jr., John D. Millett, Seymour Martin Lipset* New York: Columbia University Press, 1955 Pp. x, 326. $4 50

WILLIAM R. PABST, JR., *Washington, D. C.*

This history of the Faculty of Political Science is written in two parts. Part I, by R. Gordon Hoxie emphasizes the administrative aspects of the establishment of the School of Political Science in 1880 as an adjunct and extension to the School of Law, and its growth and development into the Faculty of the present day. Part II contains a history of each of the six existing departments, written by members of each department.

The development of statistics at Columbia reflected in this history illustrates the changing character of its subject matter as much as its institutional position. The history of statistics as a unified subject matter discipline is tersely presented in the chapter on the Department of Mathematical Statistics written by Theodore W. Anderson, Jr. In this, Anderson shows that the Department was established in 1946 as a result of the efforts of Harold Hotelling, who ironically was called to the Institute of Statistics at the University of North Carolina just before the new department was finally recognized. Hotelling initially came to the Department of Economics in 1931 to carry on the mathematical work of H. L. Moore, but in subsequent years turned from mathematical economics to concentrate on mathematical statistics. In 1938, he brought Abraham Wald to Columbia as an assistant. Wald became the first head of the newly established department in 1946.

The history of applied statistics at Columbia is contained in the accounts of four of the five other departments. In the account of the Department of History, there is reference to the work of Richmond Mayo-Smith, whose *Science of Statistics* (two volumes) was at one time widely recognized. Under the Department of Sociology, one will find the names of Franklin Giddings, Robert E. Chaddock, who was appointed Assistant Professor of Statistics in 1911, and was in 1925, president of the American Statistical Association, William F. Ogburn, also a president of the American Statistical Association, in 1931 and past editor of this *Journal*, and Frank A. Ross, also a past editor of this *Journal*. Under the Department of Economics, the history contains the names of Henry L. Moore, Wesley Mitchell, Frederick C. Mills, the latter two of whom were presidents of the American Statistical Association in 1918 and 1934, respectively. In the Department of Anthropology Franz Boas' course in statistics, termed "wholly theoretical", remained the "foundation of his teaching for forty years".

The chapter on the history of the Department of Economics written by Joseph Dorfman is especially noteworthy for the tenderness and affection with which he assesses the contribution of his predecessors and his colleagues, even to noting the "humanizing, intangible values" contributed by Mrs. Gertrude D. Stewart, secretary to the department for almost half of its seventy-five years.

Quality Control Through Statistical Methods. *Norbert Lloyd Enrick* Institute of Textile Technology: Modern Textiles Magazine Handbook No 3, 1954.

ROBERT J. HADER, *North Carolina State College*

NORBERT ENRICK, Research Statistician, Institute of Textile Technology, has prepared a generally excellent manual on quality control for the textile industry. In the past few years there has been a very rapid increase in interest in statistical quality control in this industry. The American Society for Quality Control now has a newly formed and very active Textile Section. Numerous articles on the subject have appeared in the literature, in fact, the material of this handbook originally appeared as a series of articles in *Modern Textiles Magazine*.

Most modern mills have quality control in some form, though as yet disappointingly few use the statistical tools found so useful by other industries for this purpose. Enrick's manual is intended to present these techniques—control charts, "analysis of variations," and sampling acceptance inspection—in simple straightforward manner using textile illustrations and textile language. Enough theory is given, though in non-mathematical form, so that the reader will have an appreciation of the logical foundations of the methods.

The first four chapters are concerned with frequency distributions, averages, and measures of variation. The arithmetic average, median, mode, harmonic mean, standard deviation, coefficient of variation, range, mean deviation, and average per cent variation are all included. Somehow the

"upper half mean length" has been overlooked. On the whole the discussion of the relative merits of these measures is good.

Chapters 5 and 6 deal with "Analysis of Variations" and "Process to Process Variations Analysis". The author shows how sources of variation in the process may be analyzed by techniques analogous to analysis of variance but using ranges instead of variances. This material is particularly well presented. If suffers somewhat from lack of discussion of sample sizes necessary to carry out such analyses.

Chapters 7 and 8 are devoted to control charts for averages, per cent defective, and defects. Again, insufficient attention is given to the sampling problems encountered in a quality control program in textile manufacturing. Not the least of these problems is that of providing effective and economical routine control on large batteries of machines of the same type when quality aberrations are likely on individual machines.

Chapters 9 and 10 are on "Sampling Economically and Effectively" and "Ready Made Sampling Plans". In the latter chapter sequential sampling gets heavy emphasis. A set of single sampling plans is given but passed over rather hurriedly.

There is a final short chapter on "Management Aspects of Quality Control."

All in all it is the reviewer's opinion that Enrick's manual is to be highly recommended to its intended audience.

City of Birmingham Abstract of Statistics Number 3, 1952–1954. Edited by Richard Padley, Ph D, Statistical Officer to the Corporation, and A. B. Neale, B Com Published for the Corporation by the City of Birmingham Central Statistical Office, by order of the General Purposes Committee, 1954. Pp. 139 Paper. Ten Shillings and Sixpence.

DE VER SHOLES, *Chicago Association of Commerce & Industry*

THIS compilation of statistical data for a city is interesting both for its content and its method of compilation. Few if any, American cities have Central Statistical offices which publish local data in as comprehensive a form as has been done by Birmingham, England.

The Birmingham report covers the following major statistical categories: Population; Vital Statistics; Health; Housing; New Building; the City Estates (public housing); Central Redevelopment Areas; Education; Police and Fire Services; Care of Children; Miscellaneous Social Services; Transport, Communication and Water Supply; Employment, Industry and Trade; Government; Meteorology. The data are presented in tabular form throughout, with little text material.

The data reveal, to some extent, the nature of the British way of life in which governmental information is available on many phases of urban activity for which American cities have no systematic reporting. This is especially true of the health statistics, which are in greater detail than is

generally found in American communities. On the other hand, the industrial statistical coverage, other than employment data, is almost completely lacking. The only data on industry other than employment are for size of manufacturing plants by number of employees and for factory buildings under construction.

The population statistics are in no more detail than is generally available for the United States cities. The smallest geographic units are the 38 wards, for which population is estimated annually 1946–1953. Birmingham evidentally has problems of continuity of small area boundaries similar to United States cities, as the Ward boundaries were completely revised in 1949, so that all ward statistics break at that point.

It is also interesting to note that in line with many American cities, Birmingham grew by 12 per cent in the period between 1940–1953, according to the annual estimates of population. Detroit, for example, grew 14 per cent over the same period. Birmingham had a total population of 1,118,000 in 1950; Detroit had a 1950 population of 1,850,000.

Some tables, prepared by the Central Statistical Office, are presented on the extent of annual net migration. This is generally the unknown factor in estimating population of most large United States cities. The Birmingham Central Statistical Office calculates annual net migration as the difference between the population estimated from life tables and the population estimates of the Registrar General. No statement as to the reliability of those estimates is given.

Under Chapter 2, "Vital Statistics", there are three sections: 1. Births; 2. Marriages; 3. Deaths Most of the information under this chapter is available for American cities, although much of it is unpublished.

Similarly Chapter 3, the largest chapter in the volume, containing 36 tables covering 30 pages, has brought together many statistics on different phases of health, infant mortality, infectious diseases, civic health services, much of which is available upon request in large United States cities but which few United States cities publish Chicago, for example, has available probably much more data on infant mortality than any other city in this country, and has these data broken down in 935 census tracts within the city limits The Birmingham data are for wards only, of which there are 38. However, there is very little detail on infant mortality actually published in Chicago, much less than has been presented in the Birmingham volume.

Other subjects covered in this volume are handled in much the same manner as population, vital statistics, and health statistics just discussed, although with fewer tables. Few American cities publish annual reports in which the statistical data pertaining to the city's population and industrial growth are presented. Even where such reports are made, statistical data are almost never presented in the detail found in the Birmingham volume.

One feature of the Birmingham publication which can be somewhat improved is the narrative description of the contents of the various chapters. While some pitfalls in the use of tables are pointed out, there is very little

information on the method of compilation of the statistics, especially where the data are estimated. It is very difficult, therefore, to formulate any opinion as to the validity of the estimating procedures or the reliability of the estimates themselves. A somewhat expanded explanation of each table would be extremely helpful.

The outstanding feature of the *Birmingham Abstract of Statistics* is that it is published by the city's own Central Statistical Office, which seems to be an almost unheard of adjunct to American city governments. Although American cities have the same data available, and some departments of city governments may publish similar data in their own annual reports, none of the United States' cities, as far as this reviewer is aware, maintain a Central Bureau charged with collection and publication of pertinent statistical data. It would be helpful for city governments in the United States to secure copies of the Birmingham publication as an example of what a city's Central Statistical Bureau can produce in the way of factual information helpful to business, welfare agencies, medical interests, city planners, and others who are interested in local area data.

The Redistribution of Income in Postwar Britain. *Allen Murray Cartter.* New Haven: Yale University Press, 1955. Pp. viii, 242 $5 00.

SELMA F. GOLDSMITH, *U. S. Department of Commerce*

THIS interesting book presents estimates for 1948–49 of the distribution by income brackets of the several types of central government tax liabilities in Great Britain and of the various benefits received by consumers in the form of central government expenditures. For taxable income brackets up to £500 benefits received are found to exceed tax liabilities by a substantial amount, and for higher income brackets the reverse is the case. The break-even point is somewhere between £550 and £650 of taxable income depending on which of three assumptions is used as to the distribution by income brackets of certain types of "indivisible government expenditures," (civil expenditures for general governmental operations, expenditures on the armed forces, interest on the national debt, and the government surplus.) About 80 per cent of the population is estimated to have been on the gaining end.

The greater part of the volume consists of an analysis of the figures and a description of the methodology and assumptions used in deriving them. Thus the first 6 of the 17 chapters in the book summarize the problem and the results, and the last 7 chapters present an admirably clear and detailed technical appendix on methodology. The intermediate chapters include two comparisons of the 1948–49 results, the first with the well-known estimates of the distribution of taxes and government benefits in Great Britain in 1937 developed by Tibor Barna, and the second with United States estimates for a prewar and postwar year.

The basic income distribution used by Cartter throughout the study is the frequency distribution of tax returns classified by taxable income brack-

etc., supplemented at the bottom of the income scale by an estimate of the number of "tax-family heads" below the tax-exemption limit. To taxable income in each bracket Cartter adds nontaxable types of personal income as well as "nonpersonal income" which consists mainly of undistributed corporate profits. The latter item is added to match the corresponding inclusion of corporate profits taxes in his tax liability distribution. These additions serve to increase the amount of total income in each taxable income bracket so that average incomes are sometimes above the upper limit of the bracket. Like other workers in this field, Cartter makes no attempt to shift the tax returns to income brackets comparable in definition to the income aggregates within the brackets, although he makes some adjustment for this factor in discussing his results.

Cartter's use of income tax returns as the unit of classification in his income distribution raises an interesting point of difference between the statistical work on U. S. and U. K. incomes. Most of the recent analysis of tax incidence in this country has been in terms of an income size distribution of spending units or family units (e.g., the studies by Musgrave, Tucker, Adler, and the U. S. Departments of Commerce and Labor), whereas workers on British tax incidence apparently feel that tax returns are reasonably close to a family or spending unit classification and use them as such (e.g., Cartter, Barna, Seers, and the Central Statistical Office). Perhaps this is the case, but the available figures are somewhat puzzling. Thus there were about 14½ million private households in Great Britain in 1951 according to the Census of that year as compared with the 23 million tax return units used by Cartter (including the 3 million family heads he adds at the bottom of the income scale.) If the 23 million figure approximates the number of spending units in Great Britain, then the 2 to 3 ratio of private households to spending units that obtains is far different than that found in the United States. In 1950, for example, there were 43½ million private households in this country, and, according to the Survey of Consumer Finances, some 52 million spending units, a ratio of about 5 to 6. Do the two countries really differ so much in this respect or is this another instance where comparisons are distorted by differences in definitions? At any rate this reviewer, who has struggled with the problem of converting U. S. tax returns into family units (to allow for the large number of "supplementary family earners" filing their own returns), would welcome a discussion as to why comparable adjustments are not required in the U. K. statistics.

Cartter's allocations of the various taxes among income brackets raise a number of questions that recall the lively discussion on this subject that ensued after Richard Musgrave published his estimates in the *National Tax Journal* in 1951. Taxes on corporate profits, for example, are allocated among income brackets by Cartter (as are undistributed corporate profits) on the basis of estimated holdings of shares. Musgrave, in his standard case, assumed that part of the comparable U. S. tax was shifted to consumers, part was shifted back to wage earners, and only the balance—somewhat over one-

half—was allocated on the basis of shareholdings. Cartter's methodology results in corporate profits taxes which are progressive throughout the income scale (except for the lowest bracket), whereas Musgrave found them progressive in the United States only in the income range above $7,500. While this reviewer is inclined to favor Cartter's procedure, the main point is that differences in the findings for the two countries must be examined with care.

The most interesting chapter in the book is that comparing the 1948–49 results with Tibor Barna's results for 1937. The conclusion is that the distribution of income before taxes and benefits was slightly less unequal in postwar than in prewar Britain, and that the redistribution of income, measured by comparing income minus taxes plus benefits in the two years, was more effective in reducing inequality in 1948–49 than in 1937, although not markedly so.

The other chapter on comparisons—with United States figures—is much less satisfactory. For this country Cartter uses estimates of the distribution among income brackets of spending units, income, taxes, and government benefits developed by John Adler, one of the contributors to *Fiscal Policies and the American Economy*. But Adler's U S. figures present the surprising conclusion that income inequality in the United States (before taxes and benefits) was practically the same in 1946–47 as in 1938–39 Kuznets and others working in this field have found, of course, that this was not the case; inequality as measured by the relative shares of total personal income received by the top 5 per cent of the population has since been shown by Kuznets to have declined markedly since 1939 Although Cartter notes Kuznets' findings in a footnote (Adler's figures were published in 1951 and Kuznets' in 1953), a better decision in view of the differences in the findings would have been to discard the entire comparison of the two countries, or better still to postpone such a comparison to a later book. Cartter finds that the decline in relative income inequality between the late 1930's and the late 1940's was greater in Great Britain than in the United States when he uses Adler's estimates, but that the reverse is the case when he uses Kuznets' series. Obviously the conclusion rests on the choice of figures, a point on which Cartter takes no stand.

A second study that would compare income distribution changes in the two countries would be a most interesting contribution that appears to be feasible in view of recent studies of tax incidence in this country. Attempts to estimate the distribution of benefits received among income brackets are in large part conjectural and this reviewer, like Harold M. Groves, believes them to be a fairly unprofitable occupation. But a comparative study of before-tax income distribution and of tax incidence in the United States and the United Kingdom would be of real significance and one which Cartter is well qualified to produce.

An Essay on the Economic Theory of Rank. *R. H. Tuck.* **Oxford: Basil Blackwell, 1954. Pp. 52. Seven shillings six pence.**

ROBERT M. SOLOW, *Massachusetts Institute of Technology*

FREQUENCY distributions of economic magnitudes—personal income and wealth, corporate earnings, the size of firms, the length of stretches of unemployment—nearly all tend to be highly skewed. Apart from the problems of statistical inference thus created, this fact offers a theoretical challenge. Given that human abilities are roughly Gaussian, as are most of the other physical and biological conditioning factors, what process generates out of this symmetrical raw material the gross asymmetry of the observed economic facts? This is an old problem. Pareto's[1] celebrated hyperbolic distribution curve was originally offered as a purely empirical fit, but various rationales for it have since been found, most recently by Champernowne.[2] Perhaps the commonest explanation at the present leads to the logarithmic-normal distribution as prototype The reasoning is essentially what Gibrat[3] called the law of porportional effect; incomes are generated by a sort of diffusion process except that independent random shocks have effects proportional to the displacement already achieved. The result is naturally a central limit theorem for log income J. C. Kapteyn[4] and S. D. Wicksell[5] worked along these lines years ago, and it is still being developed. H. L. Moore[6] wrote on this general problem, and Milton Friedman[7] has recently put forward a somewhat different sort of hypothesis. In contrast to the diffusion type of argument there have been some recent efforts[8] to use discrete random processes for a model.

Tuck's essay is in this general line of descent. He concentrates primarily on the distribution of firms by size (number of employees) and his reasoning about incomes is rather more casual The basic theoretical structure goes something like this: It is the nature of human organizations to be hierarchical. There is a limit to the number of subordinates one man can supervise. Hence as a firm grows in size it piles up layers of executives: one at the top supervising, say, three vice-presidents, each of whom in turn supervises three vice-vice-presidents, and so forth down to the basic operatives. Occupational rank can be measured by the number of layers between oneself and the dirty work. Now individuals enter this hierarchy and diffuse through it according to their initial training, their native ability, and their good

---

[1] Pareto, Vilfredo, *Cours d'Economie Politique*, Lausanne F Rouge, 1897, Vol. 2, Pp 299–345

[2] Champernowne, D G., "A model of income distribution," *Economic Journal*, LXIII (1953) 318–51.

[3] Gibrat, R , *Les Inégalités Économiques*, Paris. Librairie du Recueil Sirey, 1931.

[4] Kapteyn, Jacobus C , *Skew Frequency-curve in Biology and Statistics*, Groningen: Astronomical Laboratory, 1903.

[5] Wicksell, Sven Dag, *The Genetic Theory of Frequency*.

[6] Moore, Henry Ludwell, *Laws of Wages*, New York· Macmillan, (1911), 71–103.

[7] Friedman, Milton, "Choice, chance, and the personal distribution of income," *Journal of Political Economy*, LXI (August 1953), 277–90.

[8] Bernardelli, Harro, "The stability of income distributions," Sankhya, 6 (Part 4, 1943), 351–62. Solow, Robert, *The Dynamics of the Income Distribution*, unpublished thesis, 1951, Harvard University Library.

fortune. But because technologies differ among industries, the relation between rank and the attached income will also differ. In a state of equilibrium it is to be expected that the distribution of income will be about the same in all industries, else entrants would be attracted from the less-favored to the more-favored industries. But this implies that the distribution of occupational ranks will differ from industry to industry. Now the assumption is made that in any industry there will be more individuals at any given rank than can be efficiently supervised by the available people of the next highest rank. Tuck concludes (or assumes) that the excess individuals at any rank will split off and found independent firms of a size just so large that a person of that particular rank can function as chief executive. This provides employment for a chain of people of lower rank, and again the superfluity drains off into independent firms. Continuing this process to the end, a distribution of firms by size is determined. On the further assumption that in each industry log income is proportional to rank, an income distribution follows.

One gets nowhere in this business except by starting with highly simplified assumptions, and I don't mean to criticize Tuck for doing what is necessary and doing it in an original way. But the particular assumptions he has chosen do strike me as none too sturdy a foundation to support the weight of some of the strongest empirical regularities in economics. It seems somewhat far-fetched to believe that firms come into being, grow, and contract in response to the availability of people at certain occupational ranks. Nor does the picture of the capital market implicit in the theory seem to capture reality in a way that would satisfy most economists. But no star is wholly lost, and some of Tuck's ideas provide a valuable insight into the nature of inter-industrial occupational equilibrium.

The argument is supported by fits of the theoretical distribution to the observed pattern of firm sizes in the factory trades in the U. K. as of 1935, and to the iron and steel and non-ferrous metals trades in particular. The fit is impressively good, but one's awe is somewhat mitigated by the fact that the theoretical distribution contains six free parameters in addition to the sample size.

Demand for Meat. *Elmer J Working* Chicago. The Institute of Meat Packing, 1954. Pp. xi, 136 $1.00 Paper.

J. A. NORDIN, *Iowa State College*

THIS monograph is directed at prediction of the prices of beef, pork, and all meat as a composite commodity. Single equation models are used. The demand equations are fitted on the basis of the period 1922–1941, and postwar predictions are made.

Working has given us a careful and resourceful analysis of the forces affecting the sales of meat. The area of study is one in which there is considerable room for disagreement on broad methodological grounds; some such disagreement is indicated below. But though it is not to be expected that the

author's decisions on methodology will command unanimous assent, it is clear that they have enabled him to make another stimulating contribution to price analysis.

The author's principal conclusion pertains to the difference between long run and short run elasticity of the demand for meat. Price is made a function of both current consumption and past consumption, in effect. The short run price elasticity of demand is derived from the partial effect of current consumption. The long run price elasticity is derived on the assumption that current consumption equals average consumption per capita over the past ten years. The author estimates the former elasticity at 0.75, and the latter at 1.25. He suggests that restricting meat production might increase farmers' incomes in the short run but reduce them in the long run.

In this connection the author's conceptions of demand and demand elasticity are interesting. He argues that, in the case of meat, consumption is the cause of price, the quantity offered for consumption being largely determined by events before the period in which the price is set (p. 8).[1] He implies that in other problems the demand relation might be the more usual one representing quantity consumed as the effect of price.

The price elasticity of demand is written in the usual way, even though in the case of meats the numerator represents percentage change in a causing variable.

If the demand equation for the ith good involves other individual goods, they are represented by quantities rather than prices. Thus in the case of pork the demand relation indicates how pork price responds to pork quantity given the quantities of other meats.

Of course holding other prices constant is an alternative to holding other quantities constant. And recently Friedman and Bailey have discussed holding constant the consumer's indifference index.[2]

At this point the general orientation of the monograph becomes significant. The study is a study in price analysis, and relates primarily to price prediction. But it can be argued that prediction of itself is not highly important. What is important is the action-choice based on the prediction or analysis. Thus if the purpose behind the analysis is that of chosing between two price-fixing systems, and if we are concerned with the result of varying the price of pork while keeping other meat prices constant, then the usual definition of demand (in terms of response of quantity to price) is reasonable. If we are primarily concerned with consequences of output restriction, then the author's definition of demand is appropriate. If we plan to compensate a social group for changes damaging its members the Friedman-Bailey ap-

---

[1] It is not evident why he emphasizes this view, since later (p. 39) he dissociates "dependent" and "independent" from causality, and in estimating the coefficient connecting a pair of variables uses a geometric mean of the coefficients derived by minimizing the sum of squares in several directions (p. 40).

[2] Friedman, Milton, "The Marshallian demand curve," *Journal of Political Economy*, 57 (1949), 463–95, Bailey, Martin J., "The Marshallian demand curve," *Journal of Political Economy*, 62 (1954), 255–60.

proach is reasonable. Specifying the use to which the work is to be put seems indispensable.

The author reaches conclusions about both price control and output restrictions (p. 4). But the study does not appear to have been organized with the specific objective of facilitating choice among actions pertaining to either price control or output restriction. It is not clear that each methodological decision has been made on the basis of its effect on specified action problems. Of course it is hard to guess how much such a procedure would have changed the results, certainly the author's findings, such as those dealing with the differences between long and short run demand elasticities, appear to be very useful.

A second important conclusion is that "when the price level is rising, if meat prices only keep pace with the average price level of all consumer's goods, consumers will demand an increased amount of meat. Consequently, during a period of inflation and rising prices, if there is no increase in meat supplies available to consumers, meat prices, being flexible and dependent upon supply and demand in the short run as well as in the long run, will rise more rapidly than the general level of commodity prices" (p. xi). Thus in one regression analysis (p. 42) when meat consumption and deflated disposable income are held constant a one per cent rise in the price level is associated with a 1.17 per cent rise in the price of meat. This conclusion and evidence might well be examined further since it is not clear a priori why inflation should bring about such results.

A third important conclusion is that the notoriously difficult prediction of postwar prices on the basis of interwar functions is greatly facilitated by including an independent variable to represent the average of the past ten years' deflated per capita incomes. Thus long continued changes in real income are said to be more significant for meat prices than are equal real income changes of shorter duration This conclusion appears to be well founded a priori, and introducing a variable for deflated per capita disposable income for the previous ten years appears to bring about a very significant improvement in the ability of an interwar model to predict postwar prices.

Although not directly involved in the author's main conclusions several other issues are interesting. "Demand Index A" and "Demand Index B" have been used as dynamic indexes of the demand for meat. The A index is the result of dividing an index of per capital disposable income by an index of the slow-moving components of the consumers' Price Index (p. 52). In this context "dynamic" refers to a situation in which a variable's effect depends upon its rate of change or upon the length of time which has elapsed since a change occurred. Index A is said to be a combined index of disposable income and the rate of change of prices. At this point (p. 52) further discussion would have been helpful; it does not seem that index A can do what the author has designed it to do. How does an index of slow-moving prices provide an indication of the rate of change of all prices?

Index B differs from index A only by dealing with expenditure in place of

income. Substitution of index B for index A improves the predictive power of the interwar equations; the postwar peculiarity seems to be high total expenditure relative to income, rather than high meat expenditure relative to total expenditure. Yet it may not be desirable to use index B, since using it requires information we are not likely to have when we want to make predictions. Perhaps it is partly for this reason that the author regards the use of index B as a step on the way to the use of a lagged income variable.

Least squares methods are used throughout, largely on the basis of the contention that the superiority of the simultaneous equations method cannot be shown if the simultaneous equations assumptions (e.g., the assumption that there are no errors in the predetermined variables) are not met (p. 26). The only use of simultaneous equations methods appears in an appendix written by Vincent West West sets up equations for the demand for meat, supply of meat, and income. The equations are linear in the original variables. The demand equation is over-identified, and full maximum likelihood procedure is used in getting estimates of the structural coefficients. The estimates of the coefficients in the demand equation differ only slightly from the estimates of the coefficients when the same demand equation is subjected to least squares procedure. Apparently (p. 26) Working decided against using simultaneous methods partly on the basis of West's work.

West's model differs from Working's markedly: the variables used and the form of the equations are different for the two studies, for instance. It would be interesting to see the results of applying simultaneous methods to models more closely related to Working's model. If Working is right about the significance of dynamic elements, for instance, then the fact that the two basic methods give similar results for non-dynamic models may not be significant. (Of course the suggestion that there is further work to be done does not detract from the significance of what the author has done in the present monograph.)

Multicollinearity is discussed at several points in the monograph. There is a suggestion that in interwar analysis some difficulties may have been averted by the accident that real income did not vary much (p 10). If it had varied more, its potential correlation with other independent variables might have caused the kind of trouble that has shown up in attempts to predict postwar demand. The use of time as an independent variable is considered dangerous; if time is extraneous but is correlated with some of the other independent variables, then some of their influence will be attributed to time, so that later prediction will be poor (p 45). Multicollinearity involving prices, populations, and real income interwar is said to have resulted in difficulty in ascribing a reasonable influence to each variable; postwar predictions are said to suffer, since the post-war correlations among these independent variables are lower than the interwar correlations (p. 59).

Working does not introduce time as a variable in his equations. Instead, he experiments with the use of first differences as a preliminary step in some of his investigations, while recognizing the fact that first differences tend to obscure relevant long-term relations (p. 45). His arguments against includ-

ing time as an independent variable appear cogent, although considerable doubt remains about the selection of a substitute procedure.

On all the significant issues, the author's discussion should accelerate the improvement of our procedures; in this connection his treatment of lagged variables is especially noteworthy. His suggestion on the use of the data should provide significant aid to other researchers in the field. And his empirical results represent an important addition to the stock of information useful in decision-making relating to meats.

Labor Productivity in Soviet and American Industry. *Walter Galenson.* New York: Columbia University Press, 1955. Pp xiv, 273 $5.50

SIMON ROTTENBERG, *University of Chicago*

IN HIS introduction to this volume, Galenson enumerates its purposes: "to trace the development of labor productivity in a number of Soviet industries since 1928; to compare productivity in these industries with that in their U. S. counterparts; and to arrive at some general conclusions on comparative labor productivity in Soviet and American industry." The research was done as a project of the RAND corporation and draws heavily upon Soviet materials.

Statistical comparison of labor productivity, either inter-temporally or inter-spatially, is a hazardous business, because change and non-uniformity persist in occurring. Those who venture to make comparisons must sometimes introduce heroic compensatory adjustments.

In attempting this comparison of labor productivity in the United States and the USSR, Mr. Galenson encountered not only all of the conventional difficulties which attach to inter-spatial comparisons within a country; he found also the usual international comparative problem of expressing values in units of common currency and, to make matters worse, a whole range of difficulties, unique to the Soviet Union and a handful of other countries, which are associated with the meaninglessness of valuation in "markets" in which prices are administratively determined.

To escape the valuation and conversion problems, he resorted to the expression of output in terms of physical product, and this drove him to a position from which he was able to compute productivity differences between the countries for only a few industries with "fairly homogeneous products."

He covers eight manufacturing and extractive industries—coal mining, iron ore mining, crude oil and natural gas, iron and steel, machinery, cotton textile manufacturing, shoe manufacturing, and beet sugar processing. Product diversity in the machinery case compelled him to resort to a system of makeshift pricing of Soviet output and to establish value of output per worker comparisons. In all other cases, he used the ratio of physical output to labor input. Comparisons are established for the immediate pre-World War II years and something is also said about productivity trends in the Soviet Union in the decade or so preceding this and in the postwar period.

Even physical output-labor input ratios yielded results which contain errors of some magnitude because the products of "homogeneous product industries" are not at all homogeneous when they are carefully examined. Differences are sometimes intrinsic. Men's shoes are not like children's shoes and shoes made of leather are not like those made of canvas and rubber and required labor inputs vary among them. Since the product mix of "an industry" varies among countries, international productivity comparisons come to be made for different, and not the same, industries.

The construction of meaningful physical comparisons requires the discovery of tolerable equivalents which express different products in some common unit. Sometimes, equivalence is found by assumption (as when an author says, "I shall ignore this difference, because it does not seem to be of sufficient magnitude to matter") and, sometimes, it is created by corrective adjustment. Even when differences are not intrinsic, in the sense just discussed, they occur because censal criteria are different among countries.

Mr. Galenson has had to contend with both kinds of non-homogeneity. He has been explicitly conscious of the problem and has manipulated his data with care. The results show labor productivity to have been lower in the USSR than in the United States, in the late thirties, in all the compared industries. Soviet labor productivity ranged from a low of 15 per cent of American in heavy construction machinery manufacturing to a high of 58 per cent in tractor manufacture. The median value in a set of sixteen was 46 per cent.

The most valuable parts of Mr. Galenson's study are the carefully-prepared materials on physical outputs and inputs in the selected industries. Their preparation, which required the combing of Russian-language publications, makes them available, in appropriately adjusted form, to people who cannot read Russian. The least valuable parts are his attempts at explanation of observed differences. Taken all together, the book is an important addition to the literatures of the Soviet economy and of international productivity comparison.

Diagrams in Punched Card Computing. *Fred Gruenberger.* Madison, Wisconsin: University of Wisconsin Press, 1954. Pp. 18 Text, 108 wiring diagrams. $3.75 Loose-Leaf.

R. Ziola, *Dominion Bureau of Statistics*

This publication is in two parts, a very brief text followed by a collection of 108 wiring diagrams. The text deals with problems in the preparation and use of punched cards, operating principles of the IBM 602A and 604, and explanations of some of the diagrams shown in the second part of the publication. Wiring diagrams are mostly for specialized operations on the IBM 077, 402, 405, 416, 417, 513, 521, 602A, 604, and CPC Model I, with some reference to Remington Rand equipment. These include diagrams for testing machines; calculating a correlation matrix, square root, variance, factor analysis, etc.; merging, selecting, and matching of cards; and other jobs of a computing nature.

The 18-page text touches somewhat superficially on a variety of punched card and punched card equipment topics. It includes a few sentences on each of such subjects as: detecting flaws in new cards, locating punching errors, testing equipment, guarding against intermittent machine trouble and operator error, determining what not to compute on punched card equipment, and other related topics.

The collection of diagrams gives some indication of what scientific computations are possible on punched card equipment. Whether or not the machines suggested are the most efficient way of carrying out such complex computations, however, is another and more difficult question.

While the author mentions instruction, along with statistical and scientific computation and research in punched card techniques, as a function of a punched card installation, the fact that explanations of wiring diagrams are very brief or omitted entirely limits the use of this publication for training purposes. In addition the publication deals chiefly with the solution of problems of a scientific nature which require a high degree of accuracy and involve relatively few punched cards for any one problem; some of the procedures suggested do not apply to processing of punched cards generally.

The publication is in loose-leaf form and it appears that supplements, chiefly new diagrams, will be published from time to time.

Introducción a los métodos de la estadística. (Segunda parte.) *Sixto Rios.* Madrid: 1954. Pp. viii, 193–434 * Paper.

Paul R. Halmos, *University of Chicago*

Continuing in the style and at the level established in the first part of this book, the author discusses an impressively long and varied list of topics. Among them are efficient and sufficient statistics, the method of maximum likelihood, confidence intervals, tests of hypotheses, decision functions, sequential analysis, non-parametric estimation, quality control, the analysis of variance, design of experiments, and stochastic processes. There is a long appendix (41 pages) on operations research, in which topics such as the theory of queues and the Montecarlo method are mentioned. On the whole the work gives the impression of being more eclectic than selective. There are (regrettably) not so many exercises in this part of the book as in the earlier part. The mathematical level is about the same; such things as Kolmogoroff's treatment of probability via set functions receive a brief mention only, and that near the end of the book. In sum: a good bird's eye view, but likely to be frustrating as a source for learning the details of the subject.

---

* In the review of the first part of the book, in this *Journal*, 1953, p 154, the number of pages was incorrectly printed. Instead of pp. 205, it should have been pp. xiii, 192. After the customary front matter, the pagination of the second part continues that of the first.

National Income—1954 Edition—A Supplement to the Survey of Current Business. *United States Department of Commerce, Office of Business Economics.* Washington: U. S. Government Printing Office. 1954. Pp. v, 249. $1.50 Paper

Its foreword describes this volume as follows: "Since publication in 1934 of the first of a series of national income reports by the Department of Commerce, steady progress has been achieved in extending the scope of the estimates, in improving their quality, and in making them available promptly, as well as in sharpening the concepts. A principal contribution of the present report—which is closely similar in form to the 1951 National Income supplement so as to facilitate use by those familiar with that volume—is the presentation of estimates incorporating data collected in the postwar industrial and population censuses.

"In the preparation of these new estimates, opportunity was also taken to rework many of the income and product series for the entire period back to 1929 in order to reflect additional data sources and improvements in estimating techniques. A special feature is the presentation of constant-dollar gross national product in 1947 prices instead of 1939 prices, as previously used.

"The tables presented in this volume incorporate the results of the first comprehensive review of sources and methods since the initial publication of the national income statistics in the form of an economic accounting system in the 1947 National Income supplement. While the changes that have been introduced do not alter the over-all picture of the United States economy afforded by the income and product accounts, they improve the data in many detailed aspects.

"The text material in the 1951 volume also has undergone review. This resulted principally in reworking the descriptions of statistical sources and methods to accord with the new estimates and bringing up to date the summary of economic developments.

"The statistical changes have been examined for the light they shed on the reliability of the estimating techniques. This analysis confirms the adequacy of these techniques to produce reliable preliminary measures of national output and its major components on the basis of incomplete information. Revisions for some of the more detailed components, however, were substantial and underscore the need which we have repeatedly stressed for further development of the primary data sources on which the national income estimates are based.

"The present report contains all the national income statistics of the Office of Business Economics except the annual series on income by States and the distributions of family income by size classes. With these exceptions it supersedes all previously published figures, and the series contained in this volume will be kept up to date in the monthly Survey of Current Business."

W.A.W.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 272 DECEMBER 1955 *Volume 50*

STATISTICS OF THE 1954 POLIO VACCINE TRIALS*

K. A. Brownlee
University of Chicago

The *Report* on the 1954 Poliomyelitis Vaccine Trial was produced by the Poliomyelitis Vaccine Evaluation Center at the University of Michigan, directed by Thomas Francis with Robert Korns as Deputy Director and Robert Voight as Chief of Statistical Operations. Apparently the *Report* is a joint product of the staff of the Evaluation Center.

The *Report* was launched upon the world with a volume of publicity probably unprecedented for a scientific work. The *Report* was used by the National Foundation for Infantile Paralysis to push the vaccine into mass use in the spring of 1955. The fact that the vaccine then being pushed differed very importantly from that used in the 1954 trial, in that it did not contain merthiolate, was released to the public only after the fact that some of the vaccine was causing poliomyelitis could no longer be ignored. The ultimate horror came when the vaccine caused poliomyelitis not only in some of those injected but also in associates of those injected. The responsibility for these tragic events, of course, is not that of the authors of this *Report*: they were concerned solely with the evaluation of the vaccine used in the 1954 trial.

It is impossible not to be impressed with the courage of those who undertook a task of this magnitude. The *Report* lists 312 State and Local Health Officials who participated in the field trials in the U. S. and 6 others in Canada and Finland. Listed also are 54 physical therapists, 22 epidemiological intelligence officers, 28 laboratories with their principal scientists, and the 17 members of the Advisory Committee The latter, incidentally, included three who are listed in the 1954 Di-

* An invited review article on *Evaluation of 1954 Field Trial of Poliomyelitis Vaccine: Summary Report*. Poliomyelitis Vaccine Evaluation Center University of Michigan, Ann Arbor, Michigan. April 12, 1955. Pp. xiv, 51, Appendix pp. 63. No price.

rectory of the American Statistical Association. Further, several members of the staff of the U. S. Bureau of Census were on leave to the Evaluation Center, the machine tabulations were handled by the University of Michigan Tabulating Service, and the Survey Research Center of the Institute for Social Research of the University of Michigan participated. To the above must be added countless school teachers and physicians, and lastly 1,829,916 children.

The role of the Advisory Committee is not explained. Was it merely to exist and lend weight by the authority of its names? Or was it to provide answers to specific questions when asked, and otherwise be silent? Or was it to keep a collective eye on the whole proceeding and satisfy itself that everything was according to Hoyle?

The first 51 pages of the *Report* are divided into "Plan of Study" (14 pages), "Results" (44 pages), and "Summary of Estimates of Effectiveness of Vaccine" (2 pages). The remaining 61 pages are an Appendix consisting mainly of 18 tables and 11 pages of reproductions of forms.

It is not completely clear whether the staff of the Evaluation Center were responsible for the design of the field trial as well as its analysis: on page xiii we find

> The proposal was made at that time [the autumn of 1953] by the Foundation that a Center, to conduct a scientific evaluation of the effectiveness of the vaccine, if any, be established . . .

One would imagine that "to conduct a scientific evaluation" one would have to be satisfied that the design of the trial was satisfactory. However, the only authority explicitly ascribed to the Center is for analysis:

> The conditions under which a Center would function needed to be defined. It was agreed that . . . the Center would function as the executive agency of the evaluation program, and that it alone would be responsible for the analysis of all data and for any reports.

It is thus not clear, at least to this reviewer, whether being "the executive agency for the evaluation program" included the authority for the design or merely carried the responsibility for executing someone else's design. Some light may be thrown upon this question by the Public Health Service's *Technical Report on Salk Poliomyelitis Vaccine* of June, 1955, which states (page A4):

> *August 1954.* The National Foundation for Infantile Paralysis contracted with Dr. Thomas Francis, Jr., University of Michigan School of Public Health, for analysis of the data derived from the field trial.

It seems likely that this must constitute one of the largest and most expensive clinical trials ever run, and it received, or perhaps one should say obtained, many times the press publicity of any predecessor. The direct cost, the reviewer infers from newspaper reports, was of the order of $5,000,000. It is therefore somewhat upsetting to observe that the trial avoided complete disaster only by what appears to have been a late afterthought, too late, indeed, to save more than the lesser part:

> The plan of procedure announced by the National Foundation for Infantile Paralysis and its Advisory Committee was to administer vaccine to children in the second grade of school; the corresponding first and third graders would not be inoculated but would be kept under observation for the occurrence of poliomyelitis in comparison with the inoculated second graders. This has been designated the "Observed Control" study.

Apparently some one or more persons realized the total folly of this scheme and were able to do something about it:

> There was introduced, therefore, a second plan corresponding in pattern to that usually employed in scientific investigations. Children of the first, second, and third grades would be combined. One half would receive a solution of similar appearance which should have no effect on immunity to poliomyelitis. Each child would receive the same lot of material, labeled only by code, for all three inoculations Only the Evaluation Center would have the key. A single lot of each material was, so far as possible, to be used in a given area The children in the study would be observed thereafter and all reports relating to a case of poliomyelitis would be made on a concealed or blindfold basis without knowledge of the nature of the inoculum.

It is a pity that explicit credit is not given to whomever was responsible for this change. However, only 41 per cent of the trial was rescued and the remaining 59 per cent blundered along its stupid and futile path. This dichotomy of the trial persists in the *Report*: equal prominence is given throughout to data from both the "placebo areas" and "observed areas" and at times careful reading is necessary to be sure which is under discussion. For Table XI, for example, the reviewer is unable to determine whether the data refer to the placebo control areas or to the observed areas, but presumably they refer to both.

The *Report* manifests an unwillingness to come out into the open and reject the possibility of valid scientific inferences as to the efficacy of the vaccine from the "observed areas" What for example, is one to make of (page 5):

> In observed areas where only those second grade children whose parents requested participation were vaccinated, the problem of establishing the control population was more complex.

It is perfectly true to say that it is "more complex," but to indulge in understatements of this order of magnitude is to be misleading. The plain fact is that it is impossible. Why not say so? Continuing:

> After careful consideration of various alternatives, it was decided that the total first and third grade study population compared to the vaccinated second grade population would be the most critical measure that could be applied to measure the efficacy of the vaccine.

When the "most critical measure" is worthless, why allow the uninformed reader to continue with the impression that it is worth something?

The failure to face up to the implications of the confounding of vaccination status with age and other factors is manifested in the discussion of the data (page 27):

> Table 3b presents the same type of analysis of cases in the observed study areas. The significance of differences in rates in vaccinated and control populations in the major categories are also at the .001 significance level. . . . The difference between the attack rates for bulbo-spinal poliomyelitis in vaccinated and controls was significant at .01.

The unqualified attribution of these "significance levels" to vaccination rather than to the complex of factors differentiating the two populations must be condemned.

There is no need, in the pages of this *Journal,* to discuss the reasons for the need for randomization in experiments of this type, but it is perhaps worthy of note that this *Report* contains two very good illustrations of its necessity

The first example, a subsidiary study of absenteeism in Schenectady, which was a placebo area, shows very clearly that children not requesting participation in the trial are quite different from those requesting participation. The New York State Health Department investigated 24 schools involving a total study population of 4,207 children. The extent and causes of absenteeism for six weeks after inoculation were obtained for the vaccinated group, the placebo group, and also for those children for whom participation had not been requested. Thirteen hundred and sixteen (1,316) vaccinated children had a total of 2,869 days absence and 1,304 placebo children had 2,988 days absence. The total days of absence are broken down (in Table XVII) into 34 different categories, but since only child-days are given, ordinary contingency table techniques cannot be used on the data as given. There are no very obvious differences between these two, though if we remove the "nonmedical" absences (482 and 390 respectively) we get mean absences per child of 1.813 days and 1.992

days: the difference of 9.9 per cent seems, off-hand, to be a trifle large. However, the most striking feature of this table is that the 1,360 children for whom participation was not requested had a total of only 1,458 days absence, about half that of the participating group. We can make an approximate test of significance by considering the number of categories for which the mean for the participating groups exceeds the absences for the nonparticipators. If we restrict ourselves to those 24 categories for which the total absence was 10 or more child-days, the participators exceed the nonparticipators in all but two categories. While these will not be completely independent, since it will be possible for a particular child to be absent for more than one reason within a six week period, presumably this effect will be minor and there seems little doubt that the difference is genuine. Whether the reason is that the parents of the participators hold their children back from school unnecessarily or that the parents of the nonparticipating children send them to school when they should be held at home is an interesting subject for speculation. The *Report* comments (page 18) on this data: "Absenteeism in the noninoculated population did not differ significantly" (from the vaccinated and placebo populations). The report does not state what test was used: it is very difficult to believe, in the light of the foregoing discussion, that there is not a difference between the two populations

The second place where the data of the *Report* make plain the need for strict randomization is in Table 2b, where, in the placebo area, those receiving the placebo had a significantly higher rate (57 per 100,000) of paralytic polio, than those not participating in the trial (36 per 100,000). A sample survey (page 13) established that the nonparticipants belonged to lower socio-economic groups than the participants.

These two results demonstrate the futility of the work on the "observed areas."

A considerable amount of the *Report* is taken up with the results of a scoring scheme for the degree of paralysis and with the results of attempts to isolate the virus from specimens taken from patients. From the layman's point of view, probably the most interesting and convincing set of data is that which refers to the poliomyelitis cases categorized as paralytic and not-paralytic and not-polio. Since the paralytic cases are those for which there is the least difficulty in diagnosis we are probably justified in having greatest confidence in that part of the data, though some may prefer to consider only those cases confirmed by the laboratory.

It might be noted, for the benefit of the unwary, that the formula on page 62, section 3b, for testing the difference between the rates in two groups, is valid only when, in the conventional notation for the fourfold table, $a/(a+b)$ and $c/(c+d)$ are small, that is, when the relevant binomial distributions are well approximated by Poisson distributions. In the specific cases where this formula is employed in this *Report*, of course, this is the case. This formula also contains the constant $\frac{1}{2}$, presumably a correction for continuity, but this should have a negative rather than a positive sign.

The most convincing evidence for the effectiveness of the vaccine is contained in Table 2b. Confining our attention to paralytic polio, the vaccinated group had 33 cases in a population of 200,745, whereas the placebo group had 115 cases in 201,229. With the assumption that the two populations differ only in their vaccination status and that we have independent random sampling there can be little doubt as to the effectiveness of the vaccine The corresponding rates per 100,000 are 16 and 57. The "effectiveness" of the vaccine, as defined in the *Report* (page 62) is $100(1-R_1/R_2)\%$, where the R's are the rates The over-all effectiveness is thus 71.9 per cent.

It is of interest to see how far the vaccinated and placebo populations are identical except for their vaccination status. The age distributions are given in Table III and appear reasonably similar. Incidentally, in Table III, neither for the vaccinated nor for the (placebo) control do the sums for the age groups add to the totals given.

Another point for which a comparison between the vaccinated and placebo groups can be made is with regard to their initial antibody titers. It seems to have been the intention (page 3) to take blood samples from a two per cent sample of the children, and actually a 3.6 per cent sample (14,475/401,974) was taken Only a fraction of these were analyzed, however.

Table XII gives serum antibody titers for paired sera where the first blood was less than 4 units, categorized by vaccine lot and vaccination status. By "paired sera" is meant bloods from those children for whom both pre-inoculation and post-inoculation samples were available. Here we are concerned only with the titer of the pre-inoculation sample. An antibody titer of less than 4 units implies that the child is presumably low in natural defenses to polio virus. Table XII includes results from nine lots of vaccine; however, one of them, lot number 513, does not appear in Table V giving distribution of study cases of polio by vaccine lot, so this lot is omitted from further consideration here. The *Report* gives no explanation for its omission.

In Table XII we have approximately equally sized samples for vaccinated and "not vaccinated," which presumably means injected with placebo, since presumably blood could not be obtained from the nonparticipants. For example, for Lot 304 there were 472 sera from vaccinated subjects of which 109 had titers of less than 4 units for all three types of polio virus: for not-vaccinated subjects there were 408 sera of which 113 were less than 4 units. We can calculate an approximate unit normal deviate on the assumption of the normal approximation to the binomial distribution. For the eight lots considered, in the order in which they are given in Table XII, these are 1.568, 0 394, 2.154, 1.310, 1.283, −1.102, 0.423, and −0.815. If one forms an average, weighted in proportion to the square roots of the numbers of sera in the lots, one obtains 1 853, which has a single-tailed probability of 0.032

Also throwing light upon this question is the statement (page 23):

> The differences in lots are sharply marked in those [children] without preceding antibody to any type. . In the placebo area 20 6 per cent of the vaccinated and 22.9 per cent of the placebo controls were in this class.

Here the *Report* violates one of the rules of good writing, namely never to quote a percentage without reference to the sample size—we can do little more with these figures than to conclude that they lend some reinforcement to the above suspicion that the trial may be biased in favor of the vaccine by including more children with low antibodies, and hence higher likelihood of polio, in the placebo group. Alternatives to this somewhat drastic and damning conclusion might include:

(*a*) The uneven distribution of those with poor natural antibodies has occurred solely through chance, and what we observe is the 0.032 probability that will occur with this frequency Furthermore, it might be hoped or claimed, the magnitude of the discrepancy is trivial and of no practical importance.

(*b*) There is some lack of independence between individuals: we are not, perhaps, in the classical probability situation of sampling balls independently from an urn, and this lack of independence leads to α per cent points of statistics calculated on the assumption of independence occurring substantially more than α per cent of the time If this is so, of course, it will probably also apply to the statistical analysis of the polio data.

(*c*) There was some departure from pure randomization in the taking of blood samples. It is difficult to imagine that this would not also have affected the injections of vaccine and placebo.

(*d*) The blood samples were taken strictly at random, and the vaccinations performed strictly at random, but in the selection of blood samples for analysis some bias operated in the placebo group in favor of selecting bloods of low titers and this bias did not operate, or at least did not operate to the same extent, in the vaccinated group. For example, if a child was more likely to develop polio if his initial antibodies were low, and if when a child developed polio a request might be filed through the appropriate channel for its pre-injection titer to be measured, and since more children in the placebo group developed polio or polio-resembling diseases we would expect more determinations to be made on the placebo group. However, in the placebo group, only 162 children developed polio or "not polio" and since the rate of taking pre-injection blood samples was 3 6 per cent the expected number of children in this class is 5.8. In the vaccinated group there were 82 children with polio or not polio, so we would expect pre-injection blood to be available for 3.0 of them. Clearly this is not a mechanism which could account for the observed discrepancy.

Pages 22–24 give interesting two-way tables for a number of lots showing serum antibody changes for Types I, II, and III virus between the pre-vaccination and post-vaccination blood samples. It is gratifying to find not merely the marginal distributions of antibody titers but also the data for each individual child before and after vaccination On the basis of these data the *Report* categorizes lots as good, moderate, low moderate, and poor. A good lot, obviously, gives large numbers of children high antibody titers and a poor lot makes relatively little over-all change. Table V gives the breakdown of polio cases by lot arranged according to the quality of the lots. For paralytic polio, we can calculate the "effectiveness" as 74, 55, and 79 per cent for good, moderate, and low moderate and poor lots combined (the two latter are combined here as there were relatively few children with the one low moderate lot). It is rather curious that there is no evidence here that the "good" lots were any better than the "low moderate" and "poor" lots.

Of particular interest in the light of the unfortunate experiences with the subsequent use of the vaccine in 1955 are the studies the *Report* makes of the safety of the vaccine with regard both to reactions of all types and specifically to polio. From the data in the *Report* (page 21) it seems clear that the vaccine as used was innocent in 1954. It appears from other sources, however, that to obtain nonlive vaccine was far from easy. The Public Health Service *Technical Report* of June, 1955, states (page 9):

> In March 1954, soon after vaccine production for the field trial was started, live virus was detected by monkey and tissue culture tests in 4 out of 6 supposedly inactivated lots of vaccine.

and further

> In a total of 11 positive lots the testing laboratories varied in their ability to detect the virus . . .

It is rather curious that the *Evaluation Report* is completely devoid of any reference to these problems or to the methods of manufacture. In fact, the *Report* contains no references whatsoever to medical, immunological, or virological literature—the sole reference is to the paper "Statistical Methods for Poisson Processes" in the June, 1954, issue of this *Journal*.

To summarize, 59 per cent of the trial was worthless because of the lack of adequate controls. The remaining 41 per cent may be all right but contains internal evidence of bias in favor of the vaccinated. There was hope that an independent trial would be run in Great Britain under the auspices of the Medical Research Council, but this has been abandoned since they concluded that the vaccine was too dangerous. The reviewer may seem too skeptical in feeling the need for an independent confirmation of a trial run on the scale of the present one, but he would point out that gamma globulin was triumphantly proclaimed effective by the National Foundation after a similar trial, but now considerable doubts exist as to the correctness of this conclusion.

A STUDY OF INDUSTRIAL USE OF PROBABILITY STATISTICS IN THE PHYSICAL SCIENCES

JULIAN H. TOULOUSE

INTRODUCTION

IN ATTEMPTING to canvas opinions as to the use of probability statistics in industry in the physical sciences, for a panel discussion at the Montreal Meeting of the American Statistical Association, September 10,1954, we endeavored to develop a background by a survey or questionnaire sent to men and women listed in two biographical publications, and whose title or position indicated some contact with development, research, or project problems in industry. Some were listed as directors of research or development, while others were members of such groupings. In all, 440 questionnaires were sent out, which were reduced to 423 by some which were returned because of incorrect address.

THE SAMPLE

The names, themselves, were chosen by a stratified sampling, about equal in number from *American Men of Science* and *Who's Who in Engineering* This was done to give representation from both the scientific and engineering approach to possible usage of statistics. A starting page was chosen by opening each book at random and using the last digit of the page number as a starting page at the beginning of the book. Beginning with the top, left name on that page, the entries were examined in turn until a name suitable for the intent of the survey was found. Only people associated with an industrial capacity, and in some manner with development and research, were chosen, thus eliminating those in teaching, institutional, governmental, and consulting capacities covered by others on the panel. Thereafter, the same procedure was followed every twelfth page in *Who's Who in Engineering*, and every fifteenth page in *American Men of Science*, which would give something over 200 names each These selected individuals were then sent a letter of explanation and a blank questionnaire

THE QUESTIONNAIRE

The questionnaire was a return-postcard with the questions limited to what could be printed on one side. The answers were expected to be "yes" or "no" but qualified answers were often made. Not all questions were always answered, but the response was generally satisfactory, since 63 per cent of them were returned. There were 119 replies

from those of a scientific background, and 134 from engineers. One feature was the fact that many felt so strongly on the subject that letters of explanation or additional comment often accompanied the returned questionnaire.

The questionnaire was as follows:

1. Do you use *statistically* designed research plans? ______
2. Do you have a staff member trained in probability statistics? ______
3. Does he take part in the planning of experiments? ______
4. Do you base research decisions on probability analysis? ______
5. Do you use the following:

	Never	*Some*	*Much*	*Always*
Distribution Analysis	______	______	______	______
Control Charts	______	______	______	______
Tests of Significance	______	______	______	______
Correlations	______	______	______	______
Analysis of Variance	______	______	______	______
Latin or Other Squares	______	______	______	______
Other Advanced Methods	______	______	______	______

6. Do you use the above as much as you could ______, would like ______?
7. Have you a program to increase use of probability statistics?
8. If you do not use these methods, is it: cost ______, personnel ______, not applicable ______, too new ______, too formal ______, too involved ______?
9. Do you wish a summary of the answers to this questionnaire?

Only one key was used in sending out the questionnaire. To scientists a three line return address was placed on the card, and to engineers an additional line carried this author's engineering title. Thus the two fields could be kept separate as to answers. Also, no element of persuasion was used.

The questions were not only for factual answers, but also for intercomparison purposes. Question 6 offered a measure of interest, by the contrast between the two sections, where the respondent might indicate he was using the method as much as he could (with budget, personnel and other limitations) but not as much as he would like. The answer to question 9 was an important gauge of their interest in the subject, far beyond the mere desire to see what was said.

ANALYSIS

Obviously, the answers to so many questions gave a wide degree of combination and inter-comparison. I have made no attempt to go beyond the relation between pairs of questions, and submit the comparative answers as a general summary. Each returned questionnaire was copied on a McBee-Keysort card for facilitating means of analysis.

In Table 1, across the top, appear the questions which could be answered "yes" or "no," which means that questions 5 and 8 were omitted. Under each question number appears a word which should refresh the memory as to the complete question as outlined earlier in this report. Below each question number then appears on three lines the numbers of replies where the answer was "yes" (including qualified yes), or "no," or where the space on the questionnaire was not filled out. The rest of the table needs an explanation, since the data have been summarized as briefly as possible. It has a line for each of the questions, with somewhat more of a refresher as to the context of the question, and the body of the table, except for question 5, gives the proportion of "*yes*" answers to the question on the *line*, divided as to whether the answer was "yes" or "no" to the question of the column. In other words, for example, what proportion of those answering "yes" to question 1 also said "yes" to question 2, and what proportion saying "no" to question 1, said "yes" to question 2, etc., etc. In each case 100 minus the figure quoted is the proportion of "no" (or "not answered" in the case of question 9) answers. For example, the first pair of figures, under the column for question 1 appears on the line for question 2 at 85/33. This means that of the 140 who answered question 1 as "yes," meaning that they did use statistical methods, 85 per cent had a staff statistician. (100−85, or 15 per cent said no or did not answer.) This includes the qualified answers, most of which described the statistician as self-trained, or trained in short courses and the like. Of the 111 who answered question 1 as "no," 33 per cent did have a staff statistician, or a man with some knowledge of statistics—but evidently they did not use him!

One can go through the entire table on this basis and make his own interpretation as to the meanings. I have made no attempt to test the significance of nearly similar answers. The most important differences are obvious, and a few will be pointed out. Take the answers in the column for question 1, on the line for 6c and 6w, the letters of which key to the words "can" and "would like" in the complete question. Of those who answered "yes" to question 1, 48 per cent believed they

were using these methods as much as they could. Later it will be seen that personnel more than anything else was the limitation. The answer of these same people to 6w gives the important fact—only one in seven (14 per cent) was satisfied with the extent of his use of statistics —*six out of seven wanted to use the statistical method more.* It can be inferred that many limitations of budget, people, newness, and the like imposed the undesired restrictions. Those who did not use statistical methods kept the reversed trend. Only one-quarter of those answering "no" to question 1 were evidently using some form of statistics (there may be no real inconsistency because question 1 refers to research methods) as much as they could, but even more, 7 per cent were completely satisfied with the extent of such use.

A bright spot for the side of the statistician is the fact that in answering question 7, 56 per cent of those using statistics wanted to increase this use, but only 18 per cent of the non-users wanted to learn how to use statistical methods. These approximate ratios held for the first four questions, except that few (5 per cent) of the non-users were interested in design applications. Further in column 1, the answers to question 8 may be considered. Multiple answers were possible here because more than one blank space could be checked off. There was a difference in that 57 of the 79 answering question 1 as "yes," did not fill out question 8 at all, because of its phrasing, whereas only 11 of the 61 who gave a qualified "yes" and 16 of the 111 who gave a "no" answer failed to fill out this reason for not being able to use statistics more.

REASONS FOR LACK OF USE

The 22 answering "yes" to question 1 and filling out one or more blanks in question 8 gave 43 answers, and these were combined with the 50 who gave qualified answers to question 1 and gave 66 answers to question 8. The biggest hindrance was personnel, named 35 per cent of the time. Non-applicability was given 19 per cent of the time, and other answers as shown in column 1. Those who answered "no" to question 1 reversed the major reasons, naming non-applicability 38 per cent, and personnel 25 per cent of the time. And while these ratios held closely true all across in the cost and personnel lines, it was not true of those who answered question 2, where personnel and non-applicability were named about equally. In other words, if they had a *statistician* they were less likely to name personnel, and if they *did not have a statistician*, there was more reason to name personnel as a determining reason. Hence, we might conclude that one feature in industry use is the lack of trained statisticians.

EXTENT OF USE

Question 5 needs further explanation The possible answers were "never," "some," "much" or "always " These were weighted as 0, 10, 25 and 50 respectively. All of the answers were averaged according to these weights and "scores" which might indicate the relative use of the methods were thus developed. (For example: Of 140 who answered question 1 as "yes" or "qualified yes", 8 "never" used tests of significances, 68 "some," 57 "much," and 7 "always." The score would be:

$$[(8 \times 0) + (68 \times 10) + (57 \times 25) + (7 \times 50)]/140 = 17.53$$

This was recorded as "18" in the proper place) The data in Table 1 indicate how wide the difference is between users and non-users of the statistical method, but they also indicate that even the non-users have some familiarity Note how in each case the degree of usage builds up to a high for tests of significance being less for both the more basic and the more complex. The drop-off for the more complex methods, however, is precipitous with the non-users.

ENGINEERS VS SCIENTISTS

While many more inferences and interpretations can be gained from Table 1, we will leave this to others. Let us now briefly examine the answers from those primarily with a science background with those primarily from engineering, at least so far as their listing in the two biographical indexes is concerned Actually, valid differences are hard to find A few will be pointed out.

Scientists seemed more reluctant to answer question 6c, since 32 failed to answer, as compared with 2 engineers. About equal proportions answered this question "no." The difference was therefore among those who felt that they used these methods as much as they could Apparently more engineers felt that they had reached the limit of their physical restrictions or were more decisive in their opinions. On the other hand the engineers, answering question 6w, were more reluctant to commit themselves, although fewer felt that they were using these methods as much as they would like. Again differences appear in the methods used. The science group reached higher scores, indicating more use of the methods. This shows up most in the use of tests of significance To some extent engineers only exceeded the scientists in the use of control charts, probably a direct carry-over from their use of control charts in production, and in use of correlations which again, may understandably be a feature of engineering practice.

TABLE 1

SUMMARY OF 253 REPLIES FROM 423 QUESTIONNAIRES SENT OUT

	Question Number							
	1 Use	2 Stat.	3 Design	4 Analyse	6c Could	6w Like	7 Program	9 Interest
	Number of Replies to the Above							
Yes or Qualified	140	156	133	169	84	28	96	190
No	111	96	65	75	135	140	136	
Not Answered	2	1	55	9	34	85	21	63*
	Percentage answers "yes" after yes/no above							
1 Do you use?	—	75/25	87/17	78/8	48/64	68/61	80/40	62/35
2 Have a statistician?	85/33	—	98/49	75/35	63/64	71/64	85/46	67/46
3 Use him in design?	84/14	84/1	—	57/13	60/55	64/53	81/35	57/38
4 Analyse results?	94/34	80/45	91/35	—	57/76	71/76	86/57	74/46
6. Use as much as								
c. you can?	48/24	28/18	24/20	21/23	—	43/10	26/24	42/47
w would like?	14/7	14/7	14/29	14/8	20/44	—	7/15	10/44
7. Program to increase?	56/18	53/12	59/5	49/16	33/45	18/53	—	47/9
	(Multiple answers possible—Per cent is of total answers)							
8. What hinders?								
Cost	12/8	9/8	11/6	13/5	11/9	11/12	15/7	12/5
Personnel	35/28	23/36	32/30	36/28	25/36	21/40	34/31	32/21
Not Applicable	19/38	22/30	24/33	20/46	39/22	32/15	18/34	26/46
Too New	13/9	10/9	14/10	11/8	7/14	16/12	19/8	11/13
Too Formal	6/3	4/4	4/6	6/1	5/4	0/6	5/4	5/4
Too Involved	15/13	11/13	16/15	15/12	13/14	21/14	11/15	16/11
9. Want a summary?	84/36	81/34	80/38	83/43	68/18	14/12	94/37	—
	Scores after yes/no above							
5 Extent of use	(Weighted None = 0, Some = 10, Much = 25, Always = 50)							
Distributions	12/5	11/5	12/5	11/3	9/8	12/9	12/7	10/5
Charts	12/7	12/7	12/7	12/6	10/15	14/10	13/8	11/8
Significance	18/7	16/8	17/9	17/5	12/14	17/13	17/10	9/9
Correlations	16/6	14/8	15/8	14/6	13/12	15/13	15/10	13/8
Analysis of Var	14/4	12/5	14/5	13/3	8/11	11/11	14/7	11/7
Latin Square	9/1	8/3	9/2	8/2	4/7	7/6	8/4	6/4
Other	7/1	6/1	7/1	6/1	4/5	6/5	8/2	5/3

* Answer considered "no" if not checked in space provided.

QUALIFIED AND EXTENDED ANSWERS

Beyond the answers to the questions, the letters which accompanied them were of interest. Many were in explanation of the answers, and served to provide the qualification used in our interpretation. Many more were apologetic in the case of the "no" answers. Some of the

answers are quoted here because of their indications of usage and thinking. In some cases two or more excerpts appear from a single letter. The answers are grouped according to kind as shown by the subheadings that follow. Note the contradictions:

Use. "Our primary use . . . in Production Process Control." "In this field (experimental engineering) we very definitely make use of statistics and mathematics of probability." "We do make extensive use . . . developing the broad programs for extension of the system. Another area is . . . analysis of rate schedules. Another . . . reserve capacity." "On some jobs very useful in establishing facts." "Use only in production." "Used in production but not in research." "In this field (experimental engineering) we very definitely make use of statistics and mathematics of probability." "We are interpreting in the broad sense including industrial and manufacturing research rather than just fundamental research." "Multiple curvilinear graphical correlation always used." "I use where possible . . . this is not often." "I constantly have occasion to use these methods."

Programs. "Our programs . . . one of education and developing the basic concepts with the control and research groups." "Forty-five people in research and control enrolled in a course." "Self-teaching is not a satisfactory method." "We have trained a considerable number of our staff." "No specialists. Eight men with limited training." "Some of our group have had preliminary courses in statistics and have done additional reading on their own." "Are training . . . also have consultants." "Several staff members have a good working knowledge."

Interest. "All of us in the technical and scientific (staff) are fully sold on the principles." "In our research department we are at least conscious of the statistical methods." "We are very cognizant of the matter and have talked over some of the applications." "Lack of appreciation . . . various numbers suggest an experimental design . . . since the persons primarily concerned with the program do not have an appreciation . . . they carry out their programs by changing one variable at a time." "A lot depends on the basic viewpoint . . . ranges from having nothing to do with it, as too long-haired, up to a more rational view." "Personally, I am interested in these techniques, but find other personnel untrained and uninterested."

Results. "The jump from rule of thumb . . . all the way over . . . is extremely great . . . end up with some sort of practical compromise." "The use . . . (is) . . . practically nil. This is certainly not as it should be because we are primarily a research and development organization." "Some research men use the statistical approach in drawing conclusions

but conceal the fact they have done so, or soft pedal it, in writing their reports."

The statistician. "We do not have a man specially trained (MS or PhD) in statistics . . . three or four of staff studying." "In relatively few cases does the statistician have full responsibility . . . in most cases trying to work from the restricted data . . . at considerable sacrifice to the scientific niceties."

Objections. "Methods of interpreting results appear to be quite involved and time consuming." "Statistical people make their knowledge seem very mysterious and lofty." "You are probably aware that a skilled person can draw approximately the same straight line through a group of points as would be determined by the method of least squares." "He may (apply statistics) on his work *but not on mine!*" "We are civil engineers . . . no need for statistical analysis." "Not applicable in automotive research." "But not for research planning." "Probability analysis simply not applicable."

Limitations. "We use these methods, but not as much as we would like because personnel is not available" "Basically . . . no, . . . because of lack of familiarity with its use." "Higher skilled personnel probably would increase our use." "We are checking this approach currently." "In a few research studies." "Would like to use more, but higher executives need to be educated."

Summarizing, we believe that the use of statistical methods is progressing more rapidly than most of us were aware. Interest is great even among those not yet "sold" on the idea. The younger engineer or scientist is often more aware than his superior, probably because of more recent schooling and contacts. Personnel is the greatest limiting factor where the use of the method has started. Many of those not using the method do not yet realize the applicability. Cost is not considered a factor, but some deterrent is found in the formality and involvement of the methods. Generally, desire to use the methods outweighs ability or limitations. Even among those not using the methods, 36 per cent of the respondents wanted to know what answers others gave, while users wanted answers to the tune of 84 per cent. This averages 66 per cent, and about the same number (63 per cent) took the time to answer the questionnaire at all.

Industry unquestionably is "aware" of the statistical method of approaching a decision.

TIME SERIES PROBLEMS IN AERONAUTICS*

Harry Press
National Advisory Committee for Aeronautics

This paper reviews some recent applications of random process theory to problems in aeronautical engineering. A number of random type disturbances that cause aeronautical problems are described and some of their effects on the airplane indicated. Particular attention is given to the effects of atmospheric turbulence in giving rise to airplane loads and structural stresses. The general approach used in calculating airplane responses to random disturbances by means of power spectra techniques is then described As a concrete illustration, some results obtained in a study of missile behavior in rough-air flight are presented An effort is also made to indicate some unsolved statistical problems that are encountered in these applications

INTRODUCTION

Many problems in the physical and engineering sciences are concerned with the response of a dynamic system such as an airplane, a ship, or an electronic circuit to an irregular or random disturbance in time This paper is concerned with such problems in aeronautics. In the study of these problems, the statistical techniques of random process theory are beginning to play an important role A list of references of recent work in this field is given at the conclusion of this paper along with a few references to the statistical literature. It is the purpose of this paper to describe some aspects of these applications of random process theory It is hoped that the paper will serve to stimulate interest among statisticians in random process theory and encourage the development of the theory and the solution of statistical problems arising in applications

In aeronautics, a number of continuous random-type disturbances occur and affect the safety and performance of the aircraft Examples include:

(a) atmospheric turbulence
(b) fluctuations in the aerodynamic forces associated with the phenomenon of buffeting
(c) roughness of runway surfaces (both land and water)

* Presented at the 114th Annual Meeting of the American Statistical Association at Montreal, September 11, 1954.

(d) noise in electronic systems used for automatic pilots, guidance, and gunnery.

The design of aircraft for safety, adequate performance, and successful operations requires the determination and the control of the airplane behavior under the influence of these disturbances.

SOME PROBLEMS INVOLVING CONTINUOUS RANDOM DISTURBANCES

Atmospheric turbulence. In the most general form, the problem of an airplane in flight through rough air can be considered one of a complex dynamic system under the influence of a four-dimensional random velocity vector disturbance; that is, three space dimensions and a time dimension. The complete airplane response in time involves displacements and rotations in regard to three space axes as well as flexural responses of the airplane structure itself So far, only limited approaches to this problem have been attempted. As a simple example, the effect of only the vertical components of the turbulence on airplane wing load is illustrated in Fig 1. This is one of the more important aspects of the turbulence problem.

As seen in Fig. 1, the airplane is in flight with the forward velocity V.

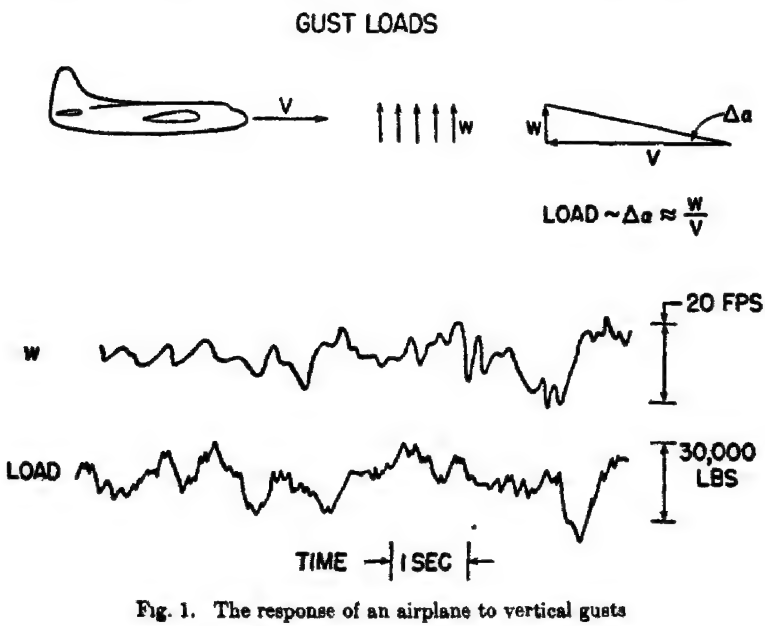


Fig. 1. The response of an airplane to vertical gusts

When the airplane encounters a vertical gust of velocity W, the airplane effectively attains an angle-of-attack change $\Delta\alpha$. For small angles, $\Delta\alpha$ is approximately equal to w/V. This angle-of-attack change in turn results in a roughly proportional increment of load on the wing, a resultant vertical acceleration of the airplane, and associated stress increments in the structural members. Atmospheric turbulence, in general, is characterized by a continuous variation in space of the air velocity. A representative variation of vertical gust velocity in space as measured by an airplane in flight is shown on the lower part of Fig. 1. These disturbances, as indicated for example by the results reported in [1], [3], and [16] appear to exhibit the principal characteristics of a random process and locally appear to approximate stationary Gaussian stochastic processes.

An airplane in flight through a gust disturbance of this type will in turn experience sizeable incremental aerodynamic wing loads such as shown by the lower curve. In addition to the vertical gust disturbance, simultaneous lateral and longitudinal velocity fluctuations will also be present and give rise to associated disturbances in the vertical motions as well as in the lateral and longitudinal motions of the airplane. These loads and motions affect the structural integrity of the airplane and may induce failure either due to overstressing or through fatigue. They may also seriously affect the stability and control of the airplane and its suitability as a bombing or gunnery platform.

Buffeting. "Buffeting" is an aerodynamic phenomenon and, as illustrated by the sketch in Fig. 2, also involves turbulent air motions. In this case, however, the turbulent motions arise from the breakdown in the smooth flow over the airplane surface. These turbulent motions in turn produce irregular and erratic aerodynamic forces on the airplane such as indicated by the traces for the differential pressure between the upper and lower wing surfaces at two points on the wing "a" and "b." These local pressure fluctuations are related in a complex fashion with pressure disturbances moving down and across the wing and undergoing continual change. Some indication of these relations can be seen from the sketches on Fig. 2 where pronounced bumps may sometimes be seen to move progressively toward the trailing edge.

These fluctuations in pressure give rise to sizeable increments of wing load and structural stresses and also affect the controllability of the airplane. An important phase of this problem is concerned with the definition of the characteristics of the pressure disturbances and their relation to the physical parameters involved, such as the airplane speed, angle of attack, and the air density. This definition would pre-

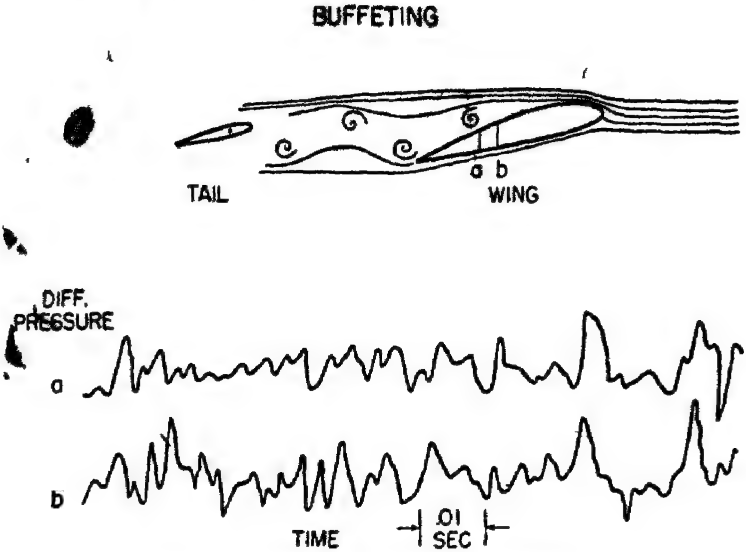


FIG 2. The phenomenon of airplane buffeting.

sumably then form the basis for the airplane response determination.

Runway roughness. The irregularities in runway surfaces affect the airplane in taxiing and during the take-off and landing operation. As shown in Fig. 3, the detail surface of a runway is not smooth, but irregular, and contains erratic variation in height. The relative height of the runway irregularities in this case has been greatly magnified for illustrative purposes. Generally, the height of these irregularities is only a few inches. With the higher speeds of modern aircraft in ground operations, these small disturbances in the runway may give rise to large airplane normal (vertical) accelerations and in some cases appear to have resulted in structural failures of aircraft components through the excitation of structural vibrations during taxiing, landing, and take-off operations. Similar difficulties are also encountered in seaplane operations in rough water.

AIRPLANE STRUCTURAL STRENGTH

One of the important effects of the foregoing disturbances on the airplane is to give rise to stresses in the structural members. In order to indicate the significance of these stresses to the structural integrity of the airplane, Fig. 4 illustrates the relation between the structural

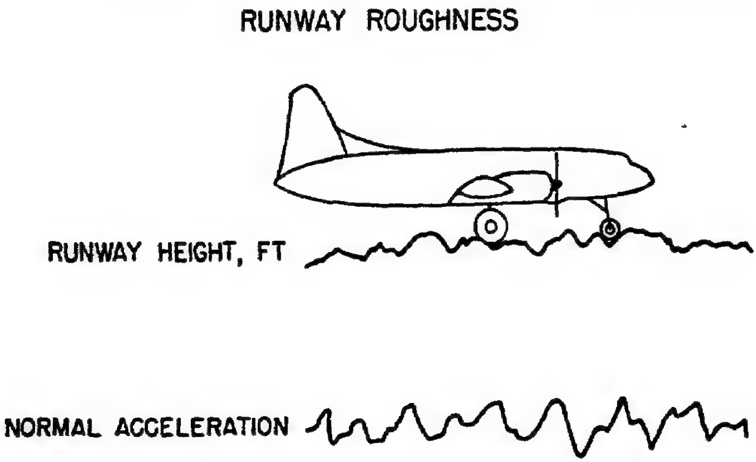


FIG. 3 The problem of an airplane landing on or taking off from a rough runway

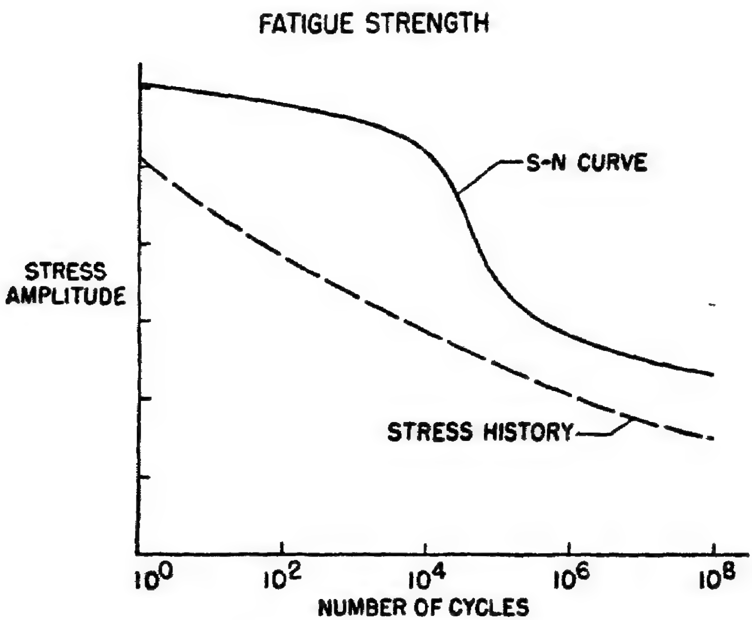


FIG. 4. Comparison of fatigue strength and stress history in operations

strength of the airplane and the stresses experienced in operations. The solid curve shown is the so-called *S-N* curve which defines the structural strength of a material or an airplane component in terms of the number of cycles of stress required for failure for various levels of stress amplitude. In the figure, one stress cycle at the highest level causes a failure by exceeding the static material strength. Also, repeated stresses at lower levels of stress can likewise cause a failure, a so-called fatigue failure.

The stress history experienced by an airplane arises from a number of causes Rough air is, in many cases, the most important source of these stresses. These stresses may be represented on the same type of plot and are shown for a given operation by the dashed line. The dashed line actually represents a frequency distribution of stresses and defines the number of stresses at the various stress amplitudes. The exact relation between a random stress history and fatigue life is not known and is an important structural problem. However, engineering rules involving concepts of cumulative damage are frequently used to estimate fatigue life from such comparisons. With continued airplane operation, this dashed line approaches the *S-N* curve at either the upper or lower level and may give rise to a structural failure by overstress or by fatigue For this reason, the prediction of the stress history is required for design purposes and, specifically, such quantities as the largest stress and the number of stress oscillations at the various stress levels are of concern The problems involved in calculating such quantities are considered in the next section

RESPONSE CALCULATIONS

The aeronautical problems considered in the foregoing discussion, in general, have the following common components: an erratic disturbance, an airplane dynamic system, and the airplane response to the random disturbance. The general approach to response calculations of this type is outlined in this section For simplicity, the discussion will be largely confined to the case of a linear system and Gaussian disturbances. For a linear system, the response at a given time $y(t)$ to an arbitrary disturbance starting from zero at $t = 0$ is given by the following relation

$$y(t) = \int_0^\infty x(t_1)A(t - t_1)dt_1 \tag{1}$$

where $x(t_1)$ is the disturbance and $A(t)$ is the response to a unit pulse at time 0 and is a form of weighting function. If $x(t)$ and $A(t)$ are ex-

plicitly known, the response may be calculated as a function of time and the characteristics of the response of concern examined in detail. This is the classical approach to the problem of response calculations for arbitrary but known disturbances.

For random disturbances, the time function is not expressible in simple explicit form. However, if the process is stationary, it may be described by an invariant property, the autocorrelation function $R(\tau)$ defined as

$$R(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t)x(t+\tau)dt \tag{2}$$

For the special case of a Gaussian disturbance, this function completely describes the process

Another, and for many purposes more useful, representation of the random disturbance is given by the Fourier transform of the autocorrelation function, the power spectral density function $\Phi(\omega)$. The autocorrelation and power spectral density functions are reciprocally related in the following manner:

$$\left.\begin{aligned} \Phi(\omega) &= \frac{2}{\pi}\int_0^\infty R(\tau)\cos\omega\tau d\tau \\ R(\tau) &= \int_0^\infty \Phi(\omega)\cos\omega\tau d\omega \end{aligned}\right\} \tag{3}$$

It is apparent from equation (3) that

$$\begin{aligned} R(o) &= \int_0^\infty \Phi(\omega)d\omega \\ &= \sigma^2 \end{aligned} \tag{4}$$

where it is assumed that $\overline{x(t)}=0$. Thus, the power spectrum can be seen to represent a frequency analysis of the variance with the element $\Phi(\omega)d\omega$ defining the contribution to the variance of the components of $x(t)$ having frequencies between ω and $\omega+d\omega$.

For a linear system, a useful and simple relation exists between the spectrum of a disturbance $\Phi_x(\omega)$ and the spectrum of the system response to the disturbance $\Phi_y(\omega)$ This relation is given by

$$\Phi_y(\omega) = \Phi_x(\omega)\left| T(i\omega)\right|^2 \tag{5}$$

where $T(i\omega)$ is the system frequency response function and describes the system response to sinusoidal disturbances of unit amplitude at the

various frequencies. The bars designate the absolute value of the complex quantity. The frequency response function $T(i\omega)$ is also simply related to the unit impulse response function $A(t)$ by the following relation

$$T(i\omega) = \int_0^\infty A(t)e^{-i\omega t}dt \tag{6}$$

Equation (5) is the basic relation between the input and output spectra and has been used widely in the calculation of system responses as well as for the determination of input spectra from measurements of system responses to random disturbances For the case of a Gaussian disturbance and a linear system, the response is likewise Gaussian and is thus completely specified by the spectrum or its Fourier transform the autocorrelation function The amplitude distribution in this case depends only on the variance.

Additional relations between the power spectrum and the time-history characteristics have been investigated in an extensive study by S. O. Rice [22] and by others. For the special case in which the disturbance is Gaussian, Rice has derived a number of relations which appear useful in aeronautical applications and are particularly significant for fatigue studies. Of these, the more important appear to be expressions for the average number of times per second, $N_c(y)$, that a given value of y is crossed with positive slope and the average number of peak values (maxima) per second, $N_p(y)$, that are above a given value of y. The average number of times per second that a given positive value of y is crossed with positive slope (or a given negative value is crossed with negative slope) is given by

$$N_c(y) = \frac{1}{2\pi}\frac{\sigma_1}{\sigma}e^{-y^2/2\sigma^2} \tag{7}$$

where σ_1 is the root mean square value of $y'(t)$, the time derivative of $y(t)$, and is related to the spectrum of $y(t)$ by the relation

$$\sigma_1 = \sqrt{\overline{[y'(t)]^2}} = \left[\int_0^\infty \omega^2\Phi(\omega)d\omega\right]^{1/2} \tag{8}$$

The expressions for the number of peaks per second that are above a given value of y (or the average number of minima or "valleys" per second that are below $-y$) is somewhat complicated but for large values of y

$$N_p(y) \approx N_c(y) \tag{9}$$

For most practical cases, it appears that this approximation is quite good for $y>2\sigma$ The total number of peak (maxima) values of y (or the number of minima) which includes the maxima for positive and for negative values of y, is given simply by

$$N_p = \frac{1}{2\pi}\frac{\sigma_2}{\sigma_1} \tag{10}$$

where

$$\sigma_2 = \sqrt{\overline{[y''(t)]^2}} = \left[\int_0^\infty \omega^4\Phi(\omega)d\omega\right]^{1/2} \tag{11}$$

and $y''(t)$ is the second derivative of $y(t)$ with respect to t It may be remarked that, at least for some cases, the loads experienced by an airplane appear to follow a simple Gaussian distribution. (See, for example, [15] and [17].) The foregoing relations between the spectrum and the peaks, equations (7) to (11), which apply to Gaussian disturbances have been tested empirically in several aeronautical applications (for example, [17]), and have been found to apply reasonably well.

It is worth noting that the response of a system to several simultaneous random disturbances such as vertical and longitudinal turbulence may also be handled in an analogous manner to that for the single disturbance For the case of two disturbances, the response $y(t)$ is given by

$$y(t) = \int_0^\infty x_1(t_1)A_1(t-t_1)dt_1 + \int_0^\infty x_2(t_1)A_2(t-t_1)dt_1 \tag{12}$$

where $x_1(t)$ and $x_2(t)$ are the two disturbances and $A_1(t)$ and $A_2(t)$ are the unit impulse responses of the system to the respective disturbances. In terms of the power spectra, the output is then given by

$$\begin{aligned}\Phi_y(\omega) &= \Phi_{x_1}(\omega)\left| T_1(i\omega)\right|^2 + \Phi_{x_2}(\omega)\left| T_2(i\omega)\right|^2 \\ &\quad + \Phi_{x_1x_2}T_1^*(i\omega)T_2(i\omega) + \Phi_{x_2x_1}T_1(i\omega)T_2^*(i\omega)\end{aligned} \tag{13}$$

where Φ_{x_1} and Φ_{x_2} are the power spectra of $x_1(t)$ and $x_2(t)$, $T_1(i\omega)$ and $T_2(i\omega)$ are the system frequency response functions for the respective disturbances (the asterisk designates the complex conjugate), and $\Phi_{x_1x_2}$ and $\Phi_{x_2x_1}$ are the cross-spectra defined by the following relations.

$$\Phi_{x_1x_2} = \frac{1}{\pi}\int_{-\infty}^\infty R_{x_1x_2}(\tau)e^{-i\omega t}d\tau \tag{14}$$

where

$$R_{x_1x_2}(\tau) = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} x_1(t)x_2(t+\tau)dt \tag{15}$$

The cross-spectrum in contrast to the power spectrum is complex with the real part providing a measure of the in-phase power sometimes called the co-power and the imaginary part providing a measure of the out-of-phase power sometimes called the quadrature power.

Equation (13) may be used directly to determine the spectrum of the output to two random disturbances For purposes of determining the spectra of the input disturbances from measurements of the output response, measurements are required, in this case, of two different responses Such applications have, for example, been made in aeronautics in the determination of two (or more) components of turbulence affecting an airplane in flight (see, for example, [16])

The preceding discussion has been confined to the simplest case of Gaussian processes. For non-Gaussian processes, the input-output relations of equations (5) and (13) still apply. However, in these cases, the variance spectrum does not provide all the information on the process that is required and recourse appears necessary to higher order spectra and co-variance functions. Theoretical relations for the determination and application of such higher order spectra and co-variance functions are, for example, given in [11]. However, because of their complexity, no numerical applications of these relations have as yet been made.

INPUT SPECTRA

From the foregoing discussion, it is apparent that output calculations require a knowledge of the input disturbance spectrum $\Phi_x(\omega)$ and the system frequency response function $T(i\omega)$ The determination of the frequency response function for a particular airplane is an engineering problem and can usually be obtained from theory by the solution of the equations of motion In many cases, experimental methods of determining $T(i\omega)$ are also available. The input spectrum usually has to be determined experimentally. As an example of the results obtained in experimental measurements of input disturbances, some measurements of the spectrum of the atmospheric turbulence will be described. Flight measurements of turbulence spectra have recently been made using airplanes and some of these results obtained from various sources are summarized in Fig 5

The ordinate of Fig 5 is the power density, the abscissa is frequency in radians per foot. The frequency argument $\Omega = 2\pi/\lambda$, where λ is the gust wavelength, is useful for airplane studies since insofar as the airplane is concerned, turbulence is essentially a spacial phenomena. The data shown, in most cases, are for the vertical component of the turbulence In one case, marked by the squares, the data are for the horizon-

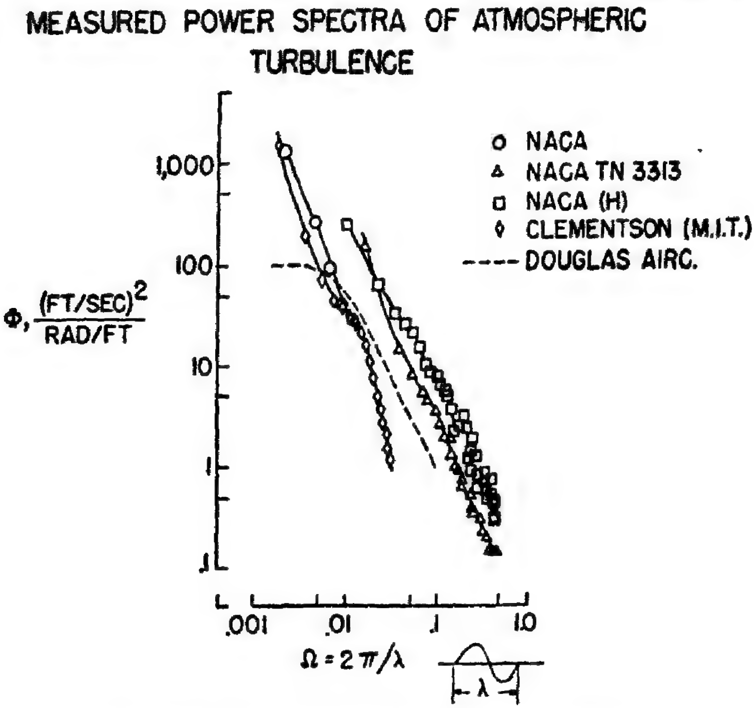


FIG. 5. Measured power spectra of atmospheric turbulence.

tal component or longitudinal turbulence. The data cover a range of gust wavelengths of from 10 to 3000 feet. which is the region that is of most concern for many airplane problems The spectra differ in intensity as indicated by the height of the spectra reflecting the variation of turbulence intensity with the weather conditions The spectra, in each case, show rapid decreases in power with increasing frequency. In fact, the spectra vary roughly as $1/\Omega^2$, a result in general agreement with theoretical results obtained in the theory of isotropic turbulence. The general shape of these spectra, with rapid decreases in power at the higher frequencies, has led to some important implications in regard to the design of airplanes for smoother flight.

For design purposes, the over-all gust experience in operations is of concern. Presumably, the over-all experience consists of various exposure times to each of the spectra of Fig 5 as well as for other spectra of turbulence associated with different weather conditions. These

spectra measurements have suggested that, as a first approximation, the spectral shape might be considered fixed and only the intensity or root-mean-square gust velocity considered to vary. Thus, the over-all gust experience may be considered a single parameter process where the parameter is the root-mean-square gust velocity. Operational data on gust loads have been used to estimate the probability distribution of this parameter, the root-mean-square gust velocity for the atmosphere. These results provide a basis for the calculation of the airplane gust response history in operations.

Spectra measurements are also being made in regard to other aeronautical problems The roughness of runways and stormy seas and the characteristics of electronic noise are a few examples in which spectra measurements are being made for purposes of design study.

AN ILLUSTRATIVE APPLICATION

In order to provide a concrete illustration of the applications of time series techniques to a particular aeronautical problem, some results obtained in a recent study [17], will be described The study to be described is concerned with the behavior and loads of a high-speed missile in rough-air flight. It was undertaken in an effort to verify the results of theoretical studies which indicated that, for the particular test conditions, missile stability changes associated with increasing forward speed would give rise to large increases in missile loads.

The missile was a tailless missile and was fired at the NACA missile test center at Wallops Island, Virginia The total missile flight lasted only about 20 seconds and yielded 2 seconds of usable record at each of eight test speeds. Inasmuch as the samples that could be obtained were necessarily small, there was some doubt that statistically significant test results could be obtained with this technique. Fig. 6 shows time histories of the missile normal (vertical) acceleration at two of the test Mach numbers which correspond to airspeeds of roughly 750 miles per hour and 600 miles per hour. Theory had predicted pronounced differences in loads for these two test conditions and this can be seen qualitatively to be the case from the relative magnitude of the fluctuations The missile vertical acceleration in both cases is seen to be somewhat irregular although the predominating motion occurs at a frequency of roughly 10 cycles per second The frequency characteristics of the records will be considered in more detail in connection with the power spectra

The distributions of the acceleration amplitude were determined from the records (readings at 0.01 second) and are shown in Fig. 7 for

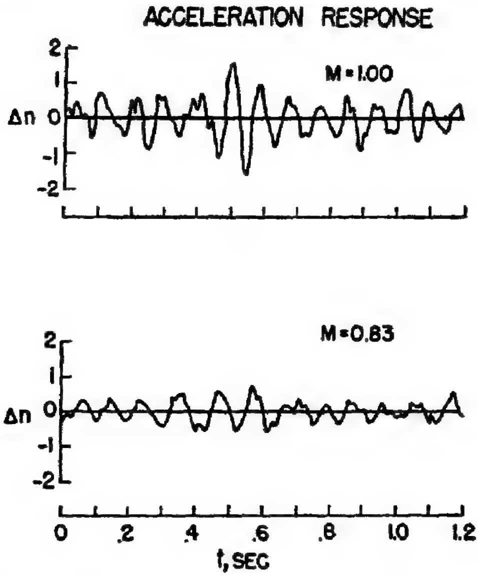


FIG. 6 Time history of missile normal acceleration in rough air at two speeds

some of the test airspeeds The figure shows the frequency distributions for several test speeds or Mach numbers. The dashed lines are fitted normal distributions The reasonable fit of the normal distribution curves suggests that the process may be considered to have a Gaussian amplitude distribution. The application of chi-square tests further tended to support this hypothesis

In Fig 8, some of the expected and measured power spectra are compared The calculated spectra for three of the test speeds are shown on the upper left and are based on airplane measurements of the turbulence over the test course and missile response theory The speeds covered here represent a speed increase of about 10 per cent from V_1 to V_3. These results, as can be seen from the relative areas or the RMS values shown, suggest rather rapid increases in power for the speed range covered. Also, the spectra move progressively to higher frequencies with increasing speed and become more peaked as a result of reduced damping. The measured spectra are shown on the lower right and were obtained by using the estimations techniques developed by John Tukey, [24]. The estimates shown are rather crude in two respects.

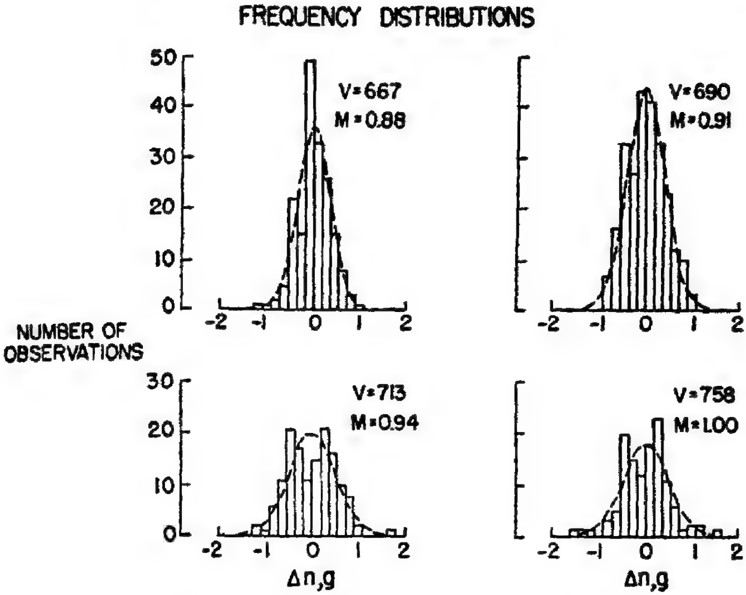


FIG 7. Frequency distributions of normal acceleration.

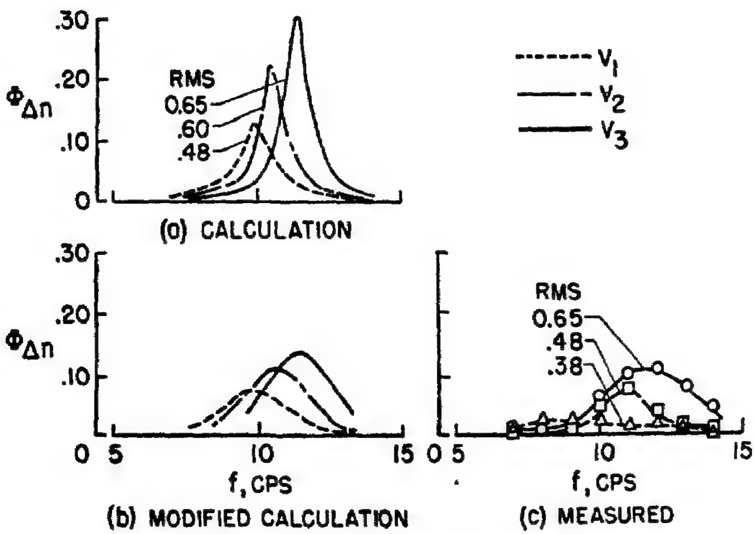


FIG. 8 Comparison of calculated and measured power spectra.

First, due to the small samples available, roughly 2 seconds for each spectrum, it was necessary to estimate average power over rather wide frequency band widths. In these cases, each point represents an average over roughly ± 1.5 cps about the values plotted Secondly, the small number of degrees of freedom for each estimate give rather poor statistical reliability for the individual estimates. The root-mean-square values are also shown and under the circumstances are considered in reasonable agreement with those calculated.

In order to compare the calculated results with those measured, the calculated spectra were averaged over frequency bands in a roughly equivalent manner to the averaging contained in the spectra measurements and the results obtained are shown on the lower left of Fig. 8. The calculated results with this modification are seen to be in good agreement with those measured and reflect the rather rapid increases in total power with speed, the progressive movement of the spectra to higher frequency, and the increasing peakedness of the spectra. These results thus appear to substantiate the theoretical expectations and have also served to establish the feasibility of the testing and analysis technique in spite of the limitations on sample size.

In order to test the reliability of theoretical results for the simple Gaussian case for the determination from the power spectrum of the number of peak amplitudes exceeding given values, a comparison of the number of accelerations counted from the test records with the number calculated from the measured spectra was made Some of the results are summarized in Fig 9 The circled points in each case represent the total number of peak accelerations (maxima and minima) obtained from the records The dashed curves represent the expected number and were calculated from the measured spectra using equations 7 to 9. The good agreement obtained here and in other such studies has indicated the feasibility of calculating these quantities by spectral techniques.

CONCLUSION

The foregoing discussion has served to describe some of the applications of random process theory to aeronautical problems. The concepts and techniques of random process theory are finding increasing use in the aeronautical sciences In many of these problems, additional developments in statistical techniques are needed to treat the problems adequately. As a conclusion, it appears appropriate to mention a few areas in which further developments would find immediate practical applications. These include:

(1) The further study of relations between spectra and time-history

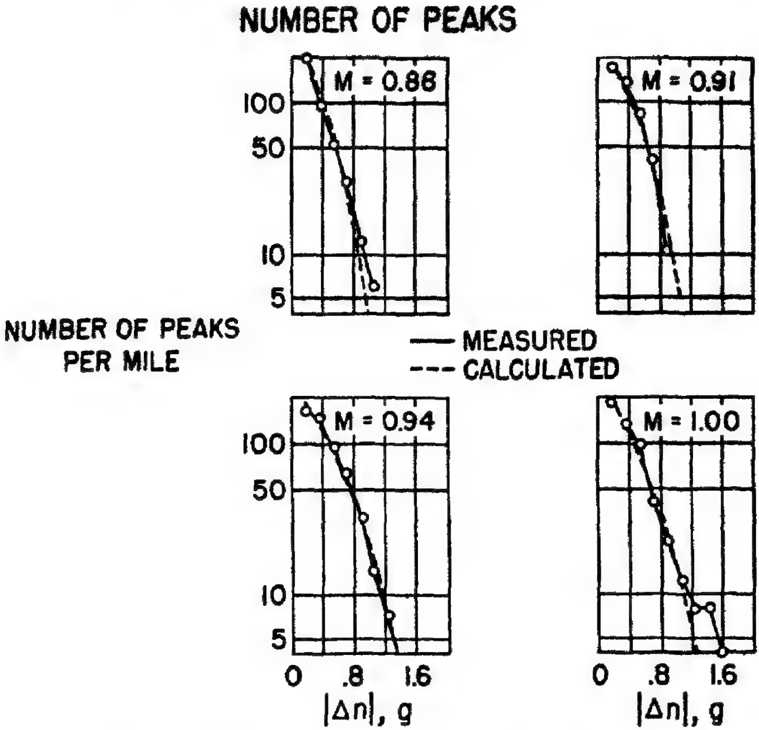


FIG 9. Comparison of calculated and measured average number of peak values per mile of flight

characteristics. For the Gaussian case, a number of relations which are proving useful have already been derived for such quantities as the expected number per unit time of crossings of given values and the expected number of peaks per unit time exceeding given values. Other quantities are also of interest for which no simple relations yet exist. For example, the expected number of ranges of given intensity, where the ranges are the amplitudes of the oscillations as measured from a trough to the succeeding peak, is a quantity of particular current interest in regard to fatigue studies In addition to the expected number for these quantities, information on the sampling variability of these quantities is also of interest

(2) The second area is that of spectra and cross-spectra estimation. The estimations of these quantities are formidable numerical tasks and

the additional development of efficient estimators is needed. The variability of some spectra estimators has been established by Tukey and others for the Gaussian case However, comparable results have not as yet been obtained for the cross-spectra case.

(3) Most of the discussion has been limited to linear systems and Gaussian processes. These two assumptions are frequently poor approximations. Unfortunately, the areas of non-Gaussian processes and non-linear operations have, because of their difficulty, so far received very little attention Further developments in these areas appear needed before many of the engineering problems can be treated adequately.

REFERENCES

Aeronautical

[1] Chilton, Robert G., *Some Measurements of Atmospheric Turbulence Obtained from Flow-Direction Vanes Mounted on an Airplane.* National Advisory Committee for Aeronautics, TN 3313, November 1954.

[2] Chippendale, George R , and Clement, Warren F., *The Statistical Study of Atmospheric Turbulence by Flight Measurements.* Report No 6822-2-T-2 (S. M Thesis), Instrumentation Laboratory, Cambridge, Mass : Massachusetts Institute of Technology, May 1951

[3] Clementson, Gerhardt C., *An Investigation of the Power Spectral Density of Atmospheric Turbulence* Report No 6445-T-31, (Sc. D. thesis), Instrumentation Laboratory, Cambridge, Mass : Massachusetts Institute of Technology May, 1950

[4] Decaulne, Paul· *Airplane Lateral Response to Statistical Gust Inputs.* Project Meteor Report (S. M thesis), Cambridge, Mass : Massachusetts Institute of Technology Guided Missles Program, May 1952.

[5] Diederich, Franklin Wolfgang: "The Response of an Airplane to Random Atmospheric Disturbances," (Ph.D thesis), California Institute of Technology, 1954

[6] Donely, Philip, *Summary of Information Relating to Gust Loads on Airplanes.* National Advisory Committee for Aeronautics, TR 997, 1950.

[7] Fung, Y C., "Statistical aspects of dynamic loads," *Journal of the Aeronautical Sciences* 20, No 5 (May 1953), pp 317–30.

[8] Houbolt, John C , *Correlation of Calculation and Flight Studies of the Effect of Wing Flexibility on Structural Response Due to Gusts.* National Advisory Committee for Aeronautics, TN 3006, August 1953

[9] Liepmann, H W., "On the application of statistical concepts to the buffeting problem," *Journal of the Aeronautical Sciences* 19, No. 12 (December 1952), pp. 793–800.

[10] Liepmann, H. W., Lauefer, J., and Liepmann, Kate, *On the Spectrum of Isotropic Turbulence.* National Advisory Committee for Aeronautics, TN 2473, November 1951.

[11] Mazelsky, Bernard, "Extension of the power spectral methods of generalized harmonic analysis to determine non-Gaussian probability functions of

random input disturbances and output responses of linear systems," *Journal of the Aeronautical Sciences* 21, No 3 (March 1954), pp. 145–53.

[12] Miles, John W , "On structural fatigue under random loading," *Journal of the Aeronautical Sciences* 21, No 11 (November 1954), pp. 753–62.

[13] Press, Harry, *The Application of the Statistical Theory of Extreme Values to Gust-Load Problems.* National Advisory Committee for Aeronautics, TR 991, 1949.

[14] Press, Harry, and Houbolt, John C , "Some applications of generalized harmonic analysis to gust loads on airplanes," *Journal of the Aeronautical Sciences* 22, No 1 (January 1954), pp 17–26.

[15] Press, Harry, and Mazelsky, Bernard, *A Study of the Application of Power Spectral Methods of Generalized Harmonic Analysis to Gust Loads on Airplanes* National Advisory Committee for Aeronautics, TR 1172, 1954.

[16] Summers, Robert A , *A Statistical Description of Large-Scale Atmospheric Turbulence.* Report T-55 (Sc D. thesis) Instrumentation Laboratory, Cambridge, Mass.: Massachusetts Institute of Technology, May 1954.

[17] Vitale, A James, Press, H , and Shufflebarger, C. C , *An Investigation of the Use of Rocket-Powered Models for Gust-Load Studies with an Application to a Tailless Swept-Wing Model at Transonic Speeds* National Advisory Committee for Aeronautics, TN 3161, June 1954

Statistical

[18] Bartlett, M. S , "On the theoretical specification and sampling properties of autocorrelated time series," *Supplement to the Journal of the Royal Statistical Society* 8, No 1 (1946), pp 27–41

[19] Bayley, G. V., and Hammersley, J M , "The effective numbers of independent observations in an autocorrelated time series," *Supplement to the Journal of the Royal Statistical Society* 8, No 2 (1946), pp 184–97.

[20] Grenander, Ulf, and Rosenblatt, Murray, "Statistical spectral analysis of time series arising from stationary stochastic processes," *Annals of Mathematical Statistics*, December 1953, pp 537–58

[21] James, Hubert M , Nichols, Nathaniel B , and Phillips, Ralph S , *Theory of Servo-Mechanisms* New York. McGraw-Hill Book Co., (1947) Chap 6.

[22] Rice, S O , "Mathematical analysis of noise " Pts I and II, *Bell System Technical Journal* 23, No 3 (July 1944), pp 282–332, Pts III and IV, 24, No. 1 (January 1945), pp. 46–156.

[23] Spetner, Lee M , "Errors in power spectra due to finite samples," *Journal of Applied Physics* (May 1954), pp 653–9

[24] Tukey, John W , "The Sampling Theory of Power Spectral Estimates," Symposium on Applications of Autocorrelation Analysis to Physical Problems, Woods Hole, Mass., 1949, Office of Naval Research, Washington, D C.

USE OF THE RANGE INSTEAD OF THE STANDARD DEVIATION

Gottfried E. Noether
Boston University

Many standard test and estimation procedures require the computation of the sample standard deviation. It is, however, often possible to replace the standard deviation by the more easily computed sample range without appreciably reducing the precision of the method. This paper discusses range methods which have been suggested in the statistical literature in connection with problems about the means and variances of one or two normal populations.

I INTRODUCTION

In a great many statistical problems it is necessary to know something about the dispersion of the population in which we are interested. The dispersion of a population is usually measured by its standard deviation σ Thus the problem arises to estimate the unknown population standard deviation from the information contained in a random sample taken from the population. The standard procedure is to estimate σ by the sample standard deviation However, from a practical point of view, the sample standard deviation has one rather serious disadvantage, the amount of labor and time required to compute it.

Often all that is needed is a quick check in order to find out whether a given hypothesis is reasonable or not We may not want to make a final decision at the moment, but only want to get a preliminary indication of what to expect in the future. In such a case it is certainly desirable to have available methods which do not require lengthy computations involving the use of a calculator.

As early as 1925, Tippett [7] showed how the range R, i e , the difference between the largest and the smallest observations in the sample, could be used as an estimate of σ. Obviously, the range is not subject to the computational disadvantage of the standard deviation. There are not many other statistics of practical importance which are as easily computed as the range

Since Tippett's early investigation an ever increasing number of papers dealing with the subject of substituting the range for the standard deviation has been published in various statistical journals. On the other hand, the methods suggested in these papers do not seem to have found widespread application except in their simplest form in statistical quality control. A reason for this is possibly the fact that few

textbooks of statistics have dealt with this question beyond the statement that the range is a rather inefficient estimate of σ if based on more than 10 or 15 observations. At the same time, the orignal contributions to the problem have often been unknown to the average user of statistical methods.

The present paper discusses certain important range methods which have been suggested in connection with problems concerning the means and variances of one or two normal distributions. The paper is divided into two parts, the first part explaining the procedures for estimating unknown parameters (primarily by means of confidence intervals) and for testing hypotheses concerning these parameters, the second part discussing more theoretical aspects. Thus, when applying range methods as short cuts for standard methods, reference to Section II is sufficient. However the reader is advised to look also at Section III in order to gain a better understanding of the methods involved, in particular of the extent to which these methods are less precise than standard methods

It will be seen that especially for problems involving means the loss in precision is surprisingly small and for many practical purposes is more than compensated for by the greater speed and simplicity of working with ranges Thus the use of range methods, instead of the corresponding methods involving the computation of the sample standard deviation, seems to be justified in many situations when it is not absolutely necessary that the last bit of information be extracted from the available data.

II APPLICATIONS

2.1. *Notation and General Remarks.* The following notation is used in the formulas of Section II:

μ = mean of underlying normal distribution
σ = standard deviation of underlying normal distribution
N = sample size
m = number of subsamples
n = size of subsamples
$\bar{x} = \dfrac{1}{mn}\sum_{i=1}^{mn} x_i$ = mean
R_i = range of the ith subsample, $i=1, \cdots, m$
$\bar{R} = (R_1 + \cdots + R_m)/m$ = mean range
a_N = constant given in Table 1
$S_R = a_N(R_1 + \cdots + R_m) = a_N\Sigma$

$$G_1 = \frac{|\bar{x} - \mu|}{\bar{R}}$$

$$G_2 = \frac{|\bar{x}_1 - \bar{x}_2|}{\bar{R}}$$

$g_{i\alpha}$, $i=1, 2$, = critical value of G_i at significance level α
h_l = factor for lower confidence limit of σ
h_u = factor for upper confidence limit of σ.

The methods discussed in this paper require that if the available sample contains 12 or more observations the original sample be subdivided into several subsamples of *equal* size, omitting some observations if necessary. The number and size of subsamples will be denoted by m and n, respectively. Let N be the total number of observations in the original sample. Table 1 shows how m and n should be chosen for $N=2(1)100$ Of course, if N is at the disposal of the statistician, it is best to choose it in such a way as to make the omission of any observations in the subsequent analysis unnecessary. As will be seen in Section III, sample sizes which are multiples of 8 are especially suitable when using range methods, though sample sizes which are multiples of 7 or 9 are almost as good.

Whatever the number and size of the subsamples, it is important that they be chosen in a random fashion. Thus if the observations are given in a random order we might simply take the first n observations as the first subsample, the next n observations as the second subsample, etc However, if the observations have been arranged in a systematic fashion, we must select n observations at random from among all N observations to form the first subsample, then select another n observations from among the remaining N-n observations to form the second subsample, etc.

2 2. *Problems Involving Means.* Problems involving means can be solved with the least amount of computational work by making use of two tables computed by Jackson and Ross [3]. These tables give critical values, $g_{i\alpha}$, say, $i=1, 2$, for significance levels $\alpha=.10$, $.05$, and $.01$ of the two statistics

$$G_1 = \frac{|\bar{x} - \mu|}{\bar{R}} \quad \text{and} \quad G_2 = \frac{|\bar{x}_1 - \bar{x}_2|}{\bar{R}}$$

which can be used to test hypotheses of the type $\mu=\mu_0$ and $\mu_1=\mu_2$, respectively. We shall describe now how these tables can be used to find

two-sided confidence intervals having confidence coefficient $1-\alpha$ for the mean of a single normal distribution and the difference of the means of two normal distributions One-sided confidence intervals having confidence coefficient $1-\alpha/2$ are obtained by using only one of the limits of the corresponding two-sided confidence intervals

A confidence interval for μ is given by

$$\bar{x} - g_{1\alpha}\overline{R} < \mu < \bar{x} + g_{1\alpha}\overline{R}.$$

Here $\overline{R}$ is the mean range of the m subsamples of size n obtained from the original sample of size N, $g_{1\alpha}$ is the critical value taken from Table I [3]; $\bar{x}$ is the mean of the mn observations used in computing $\overline{R}$

Example 1a: An I Q test has been given to 25 persons with the following results. 92, 90, 124, 110, 91, 118, 98, 107 | 128, 108, 112, 104, 100, 103, 105, 97 | 105, 99, 92, 101, 99, 113, 111, 104 | 109. To find a confidence interval for the mean score having confidence coefficient .95

From Table 1 we find that for $N=25$ we should use 3 subsamples of size 8, which means that we have to omit 1 observation Assuming that our observations are written down in random order, either as a consequence of the method by which they were obtained or by rearrangement with the help of random numbers as, e.g , The RAND Corporation's *A Million Random Digits*, we divide the total sample into subsamples as indicated by the vertical bars. We easily find $R_1=124-90=34$, $R_2=128-97=31$, $R_3=113-92=21$, so that $\overline{R}=(34+31+21)/3=28\ 7$. Also, $\bar{x}=2511/24=104.6$ Entering Table I [3] with $m=3$, $n=8$, $\alpha=.05$, we find $g_{1,\ 05}=15$ Thus the confidence interval for μ is given by

$$104\ 6 - .15 \times 28\ 7 < \mu < 104\ 6 + .15 \times 28\ 7$$

or

$$100.3 < \mu < 108\ 9.$$

The corresponding confidence interval based on Student's t is given by

$$100.8 < \mu < 108\ 8.$$

To test the hypothesis $\mu=\mu_0$ against the alternative $\mu\neq\mu_0$ we can use the above confidence interval, or, if we prefer, the statistic G_1 as indicated in [3]

One-sided tests at significance level $\alpha/2$ are carried out by comparing the statistic

$$G_1{}^1 = \frac{\bar{x} - \mu}{\overline{R}}$$

with $\pm g_{1\alpha}$ depending on whether the alternative is $\mu > \mu_0$ or $\mu < \mu_0$.

Example 1b: Using the data of *Example 1a*, to test the hypothesis that $\mu \geqq 108$ against the alternative that $\mu < 108$.

$$G_1{}^1 = \frac{104.6 - 108}{28.7} = -.12.$$

Since $-.12 < -.10$, the .05-significance point for the one-sided test, we reject our hypothesis at the .05 level.

The use of the statistic G_2 in comparing the means of two populations is complicated by the necessity of dividing both samples into subsamples of *equal* size n. Once this has been achieved, a confidence interval for $\mu_1 - \mu_2$ having confidence coefficient $1-\alpha$ is given by

$$(\bar{x}_1 - \bar{x}_2) - g_{2\alpha}\bar{R} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + g_{2\alpha}\bar{R}.$$

Here $\bar{R}$ is the mean range of the $m_1 + m_2$ subsamples of size n; $g_{2\alpha}$ is the critical value taken from Table II [3]; $\bar{x}_1$ and $\bar{x}_2$ are the respective means of the observations used in computing $\bar{R}$

If the original two samples are of equal size, $N_1 = N_2 = N$, the values of n and $m_1 = m_2 = m$ can be determined from Table 1. If $N_1 \neq N_2$, the division into subsamples suggested in Table 1 cannot be used unless we should have $n_1 = n_2$ as, e g., for $N_1 = 38$, $N_2 = 46$ for which $n_1 = n_2 = 9$ If according to Table 1 $n_1 \neq n_2$, as a general rule, it is best to choose as n that number between 6 and 10 which requires the smallest total number of observations to be omitted. However subsamples of size 5 or 11 should also be considered if their use leads to a considerable reduction in the number of observations to be omitted, as, e.g., for $N_1 = 15$, $N_2 = 20$ or $N_1 = 22$, $N_2 = 34$ Of course, if at all possible, the simplest solution is to choose N_1 and N_2 in such a way that both are multiples of the same number n, with n preferably 8 or a number close to 8.

Example 2a. Two kinds of rations fed to test animals for a given period of time produce the following weight gains.

Ration 1:

42, 38, 34, 33, 37, 35 | 35, 38, 33, 30, 36, 39 | 37

Ration 2:

33, 34, 33, 30, 36, 32 | 28, 32, 30, 36, 33, 35 | 31.

To find a confidence interval for the difference in weight gain having confidence coefficient .99.

Since $N_1 = N_2 = 13$, we find from Table 1 that each sample should be divided into 2 subsamples of size 6, omitting thus 1 observation from each sample. Using the subdivision as indicated by the vertical bars, we find $\bar{R} = (9+9+6+8)/4 = 8$. Also, $\bar{x}_1 = 430/12 = 35.8$, $\bar{x}_2 = 392/12 = 32.7$, $\bar{x}_1 - \bar{x}_2 = 3.1$. Thus the confidence interval becomes

$$3.1 - .46 \times 8 < \mu_1 - \mu_2 < 3.1 + .46 \times 8$$

or

$$-.6 < \mu_1 - \mu_2 < 6.8.$$

Tests of hypotheses concerning the difference $\mu_1 - \mu_2$ are carried out either by referring to the corresponding confidence interval or by using the statistic G_2 directly.

Example 2b: For the data of *Example 2a*,

$$G_2 = \frac{3.1}{8} = .39$$

which according to Table II [3] is significant at the .05 level, but not at the .01 level. The second statement follows already from the confidence interval given above

One-sided tests at significance level $\alpha/2$ are carried out by comparing

$$G_2{}^1 = \frac{\bar{x}_1 - \bar{x}_2}{\bar{R}}$$

with $\pm g_{2\alpha}$

Above methods for comparing the means of two normal populations can be used only if it is possible to assume that the two populations have equal variance.

2.3. *Problems Involving the Standard Deviation.* An unbiased estimate of σ is given by

$$S_R = a_N \Sigma$$

where a_N is given in Table 1 and $\Sigma = R_1 + R_2 + \cdots + R_m$, R_i, $i = 1, 2, \cdots, m$, being again the range of the ith subsample.

A confidence interval is easily found with the help of S_R

$$h_l S_R < \sigma < h_u S_R$$

where h_l and h_u are the factors tabulated in Table 1 for confidence coefficients .90 and .98.

TABLE 1

SUBSAMPLE DETERMINATION AND ESTIMATION CONSTANTS FOR σ

Sample size N	Number of sub-samples m	Size of sub-samples n	Coefficient for unbiased estimate of σ a_N	Approximate factors for confidence limits of σ: Conf. coeff. 0 90 h_l	h_u	Conf. coeff. 0.98 h_l	h_u
2	1	2	.886	41	12 7	.31	63 7
3	1	3	.591	51	3.9	41	8.9
4	1	4	486	.57	2.7	.47	4.8
5	1	5	430	.60	2 3	.51	3 5
6	1	6	.395 −	63	2 0	53	2 9
7	1	7	.370	65	1 88	.55	2 6
8	1	8	351	66	1.78	57	2.4
9	1	9	337	68	1 71	.58	2 2
10	1	10	.325 −	69	1 66	60	2 1
11	1	11	.315 +	.70	1.61	61	2.0
12–13	2	6	197	.71	1 57	.62	1.94
14–15	2	7	.185 −	72	1.50	.64	1 80
16–17	2	8	.176	.74	1 45	66	1.70
18–19	2	9	168	75	1 43	68	1.67
20	2	10	162	.76	1 41	69	1 65
21–23	3	7	123	77	1 39	70	1 62
24–26	3	8	117	78	1 36	.71	1 56
27	3	9	.112	.79	1 34	.72	1.52
28–29	4	7	0924	79	1.33	73	1 51
30–31	3	10	108	80	1 31	74	1 49
32–34	4	8	.0878	80	1.30	74	1 47
35	5	7	.0740	.81	1 28	.75	1.44
36–39	4	9	0842	.81	1.28	75	1 43
40–41	5	8	.0702	82	1 27	.76	1.41
42–44	6	7	.0616	.83	1 26	.77	1 39
45–47	5	9	0673	.83	1 25	77	1 37
48	6	8	.0585 +	.83	1 23	.78	1 35
49	7	7	.0528	83	1 23	.78	1.35
50–53	5	10	.0650	.83	1.22	.78	1 34
54–55	6	9	.0561	.84	1 22	.79	1 33
56–59	7	8	0502	.84	1.21	79	1.32

TABLE 1—(*continued*)

Sample size N	Number of sub-samples m	Size of sub-samples n	Coefficient for unbiased estimate of σ a_N	Approximate factors for confidence limits of σ — Conf. coeff 0 90 h_l	h_u	Conf. coeff 0.98 h_l	h_u
60–62	6	10	.0542	85	1.20	.80	1 31
63	9	7	0411	85	1 19	.80	1 30
64–65	8	8	.0439	.85	1 19	80	1 30
66–69	11	6	0359	85	1 19	80	1 30
70–71	10	7	0370	86	1 19	81	1.29
72–76	9	8	0390	86	1 18	81	1.28
77	11	7	0336	86	1 18	.82	1.27
78–79	13	6	0304	86	1 18	82	1 27
80	10	8	0351	87	1 17	82	1 26
81–83	9	9	0374	87	1 17	82	1.26
84–87	12	7	0308	87	1 17	82	1 26
88–89	11	8	0319	88	1 17	.83	1.25
90	10	9	0337	88	1 16	.83	1 25
91–95	13	7	0284	.88	1 16	83	1 24
96–97	12	8	0293	88	1 16	83	1.23
98	14	7	0264	88	1 15	.83	1 23
99	11	9	0306	88	1 15	83	1 23
100	10	10	0325 –	88	1 15	83	1 22

Example 3a: Suppose that for a set of 75 observations we have found $\Sigma = 295$, Σ being based on 9 subsamples of size 8 omitting 3 observations as indicated in Table 1. Then $S_R = .029 \times 295 = 11.5$. The confidence interval with confidence coefficient .98 becomes

$$81 \times 11.5 < \sigma < 1.28 \times 11\ 5$$

or

$$9.3 < \sigma < 14\ 7$$

The two end points of the confidence interval given above can also be obtained graphically The equations $\sigma = h_l S_R$ and $\sigma = h_u S_R$ represent two straight lines through the origin in the (S_R, σ)-plane Having observed a particular value S_R, we draw a line parallel to the σ-axis at distance S_R This line intersects the other two lines in the end points of

the confidence interval. This graphical method is certainly advisable if repeated confidence intervals based on samples of the same size N are to be found.

As usual, tests of hypotheses concerning σ can be carried out by checking whether the hypothetical value of σ lies in the corresponding confidence interval or not. Alternatively, the percentage points of S_R/σ tabulated by Grubbs and Weaver [2], pp 233–4, can be used.

Example 3b: Using the information of *Example 3a*, to test the hypothesis that $\sigma \leqq 10$ against the alternative that $\sigma > 10$. $S_R/\sigma_0 = 11.5/10 = 1.15$. From Grubbs and Weaver's table for 9 subsamples of size 8 we find the (one-sided) .05 significance value as 1 16. Thus we do not have sufficient evidence to reject our hypothesis at the .05 level.

Since in testing the hypothesis $\mu_1 = \mu_2$ it is necessary to make the assumption that $\sigma_1 = \sigma_2$, it would be desirable to have a quick range test for ascertaining whether this assumption is reasonable or not. Unfortunately, tables for such a test exist only for the case when the number of observations in each sample does not surpass 10.

The test consists in forming the ratio of the two ranges putting the larger range in the numerator. The null hypothesis of equal variance is rejected if the observed value of the range ratio is too large Critical values for the range ratio are given in [1], pp 317–8, Table 8d. In using the table, n_1 refers always to the number of observations on which the range in the numerator is based. As in the case of the corresponding F-test, the significance level of this test is twice the probability level indicated in the table.

III. DISCUSSION

We shall now discuss the theoretical basis and justification of the procedures given in Section II.

3.1. *Determination of Size of Subsamples.* Let $R = R_{(n)}$ stand for the range of a sample of size n. It is easily shown that there exist constants d_n depending on n but not on σ (nor the mean μ of the underlying normal distribution) such that the expected value of R is $d_n\sigma$,

$$E(R) = d_n\sigma. \tag{1}$$

thus d_n is the expected value of the range of a random sample of size n from a normal distribution with variance 1. Numerical values of d_n were computed by Tippett [7] for values of n between 2 and 1000.

We can use (1) to find an unbiased estimate of σ, namely R/d_n. For small values of n, this estimate is only slightly less accurate than that

based on the sample standard deviation s, where $s=[\Sigma(x-\bar{x})^2/(n-1)]^{1/2}$ However as n increases beyond 12 or 15, say, the loss in accuracy soon becomes considerable. There is also another reason which will be discussed at a later point why we should not use ranges based on samples of too large a size. This means that for such values of n our procedure has to be modified somewhat. Instead of computing the range from the entire sample, it is now necessary to divide the given sample into a number of subsamples and then compute a weighted mean of the various subsample ranges.

Let then N denote the total number of observations available. For $N=2(1)100$, Grubbs and Weaver [2] have determined the particular subdivision of the total sample (and corresponding weighting factors) which produce the unbiased estimate of σ having minimum variance. Actually the method of Grubbs and Weaver can be further simplified. As an example, take the case when $N=58$. According to [2] we should divide the given sample into 5 subsamples of size 8 and 2 subsamples of size 9. The sum of the ranges of the 5 subsamples of size 8 is multiplied by 0485. To this is added the sum of the 2 ranges of the subsamples of size 9 multiplied by 0521 The standard deviation of this estimate is $.1070\sigma$. We can save one of the two multiplications by omitting 2 of the observations and dividing the remaining 56 observations into 7 subsamples of size 8. The estimate of σ based on this method has standard deviation 1088σ, which is not very much larger than that of the earlier estimate. This latter method certainly has the advantage of greater simplicity and symmetry without unduly increasing the standard deviation of the resulting estimate. In general, by omitting some of the observations if necessary, it is always possible to divide the remaining observations into subsamples of equal size n.

Given a sample of size N, the question arises how n should be chosen. A good procedure is to choose for n that value between 6 and 10 which requires the fewest number of observations to be omitted. If for a given N this can be done in more than one way, subsamples of size 8 are preferable to those of size 7 or 9, which in turn are preferable to subsamples of size 10 or 6 This rule has been used in the construction of Table 1 for $N=2(1)100$.

The division into equal subsamples given in Table 1 is not always the one leading to the smallest possible standard deviation of a range estimate based on equal subsamples. Thus for $N=26$, a division into 2 groups of size 13 would result in an estimate having standard deviation $.163\sigma$ instead of $.166\sigma$ for the estimate based on 3 groups of size 8, omitting 2 observations as suggested by Table 1. There are a few addi-

tional cases. The difference in standard deviation is always negligible. The reason why the division into subsamples given in Table 1 is preferable to others giving possibly a smaller standard deviation is that, except for $N=15$, such other subdivisions correspond to subsample sizes $n>10$ Now throughout this paper it is always assumed that the underlying distribution is strictly normal. However, in practice, slight deviations from normality, especially in the tails of the distribution, usually occur. Such deviations from normality are apt to influence ranges based on a large number of observations much more than those based on a small number of observations. It is, therefore, advisable to keep the subsample size relatively small.

3.2. *Test and Estimation Procedures for σ.* Given a sample of size N, it is divided into m subsamples of size n as indicated in the previous section omitting some of the observations if necessary Let $R_1, \cdots, R_m$ denote the ranges of these m subsamples By (1), an unbiased estimate of σ is given by

$$S_R = \frac{1}{md_n}(R_1 + \cdots + R_m). \tag{2}$$

If we set $1/md_n = a_N$, we get the unbiased estimate of σ given in Section 2.3.

It follows immediately from (2) that the variance of S_R is

$$\text{var}(S_R) = \sigma^2\sigma_n^2/m \tag{3}$$

where σ_n^2 is the variance of $R_{(n)}/d_n$, $R_{(n)}$ being this time the range of a random sample of size n from a normal distribution with variance 1

For the convenience of the reader, Table 2 below gives the values of d_n, $1/d_n$, σ_n^2, and σ_n for $n=6(1)10$. These values can be used, e.g , when dealing with the case $N>100$.

As already mentioned in Section 2.3, Grubbs and Weaver [2], pp

TABLE 2

RANGE CONSTANTS

n	d_n	$1/d_n$	σ_n^2	σ_n
6	2.534	.395	1120	.335
7	2.704	.370	.0949	308
8	2.847	351	.0829	.288
9	2 970	337	.0740	272
10	3.078	.325	.0671	.259

223–4, have tabulated approximate percentage points of the distribution of S_R/σ (Grubbs and Weaver use the notation $\bar{R}_n/d_n$ for S_R) which are surpassed with probability .995, .990, .950, .050, .010, .005 for $m=1(1)15$ and $n=6(1)10$. The values h_l and h_u in Table 1 are simply the reciprocals (after some necessary smoothing) of these percentage points. The tables by Grubbs and Weaver also give the values of σ_n^2/m.

How accurate are the range procedures discussed in 2.3 in comparison with the standard methods based on the sample standard deviation s? We can compare the two estimates of σ by computing their relative efficiency e, i e , the ratio of their variances We can interpret $1-e$ as the proportion of observations which are "wasted" by using the less efficient estimate, i e., the estimate having the larger variance. In our case, the quantity e can also be used in comparing the two test procedures based on S_R and s, respectively.

If N is sufficiently large, we have to a good approximation var (s) $=\sigma^2/2(N-1)$. Thus by (3)

$$e = \frac{\dfrac{1}{2(N-1)}}{\dfrac{\sigma_n^2}{m}} = \frac{\dfrac{1}{2(mn+l-1)}}{\dfrac{\sigma_n^2}{m}} = \frac{1}{2n\sigma_n^2\left(1+\dfrac{l-1}{mn}\right)}$$

where l is the number of observations omitted in computing S_R Since for most values of N, l is either 0, 1, or 2, we get a good idea of the efficiency of the estimate S_R by computing

$$e' = \frac{1}{2n\sigma_n^2}\,.$$

We get the following table:

n	6	7	8	9	10
e'	.744	.752	.754	.751	.745

This table bears out our earlier assertion that when using range methods subsamples of size 8 are best with subsamples of size 7 and 9 being almost as good. It further shows that the relative efficiency of S_R is approximately .75 at least for sufficiently large samples From a theoretical point of view, this means that we are "wasting" one out of every four observations by using S_R instead of s. But let us look at the question from a somewhat different point of view.

We shall assume first that we are interested in the estimation problem. It seems to this writer that from a practical point of view, instead of looking at the efficiency e, we get a better idea of the loss in accuracy by comparing the average length L_R of the confidence interval for σ based on S_R with the average length L_s of the confidence interval based on s. From the formula for the confidence interval given in Section 2.3 we easily find

$$L_R = S_R(h_u - h_l),$$

and similarly,

$$L_s = s(h_u' - h_l')$$

where h_u' and h_l' are the corresponding factors based on the s/σ distribution. Thus $E(L_R) = \sigma(h_u - h_l)$ and $E(L_s) = c\sigma(h_u' - h_l')$ where c is defined by $E(s) = c\sigma$ [1]

The following is a table of approximate values of $E(L_R)/E(L_s)$ for various values of N and confidence coefficient .90:

N	25	50	75	100
$\dfrac{E(L_R)}{E(L_s)}$	1.17	1 16	1.15	1.15

Thus the average length of the confidence interval based on S_R is about 15–17 per cent longer than that based on s.

Let us now take the problem of testing a hypothesis concerning σ, say, $\sigma = \sigma_0$ against the alternative $\sigma = \sigma_1 > \sigma_0$. In considering a test of this hypothesis we should certainly ask the following question How much larger than σ_0 must σ_1 be so that we can be reasonably certain to reject our hypothesis when the alternative is true? In other words, if we set $\sigma_1 = \lambda\sigma_0$, for what value of λ do we have probability $1-\beta$, where β is some small number, of rejecting the hypothesis tested. It is well-known that $\lambda_s = \chi_\alpha | \chi_{1-\beta}$ where χ_P^2 is the $(1-P)$-percentile of the Chi-square distribution with $N-1$ degrees of freedom Similarly, it can be shown that λ_R is the ratio of the corresponding percentiles of the S_R/σ distribution. For $\alpha = \beta = .05$, we find the following table of approximate values of λ_R/λ_s:

N	25	50	75	100
$\dfrac{\lambda_R}{\lambda_s}$	1.08	1.05	1.04	1.04

[1] In the customary quality control notation, $c = [N/(N-1)]^{1/2}c_2$ where $c_2 = (2/N)^{1/2}\Gamma(N/2)/\Gamma(N/2-1/2)$.

indicating that σ_1 has to be about 4–8 per cent larger when using S_R as the test statistic than when using s.

There is one other point which should be mentioned. It follows from an investigation by King [4] that the loss in efficiency discussed earlier is primarily due to the fact that our range methods require the subdivision of the original sample into m subsamples. This produces a loss of $m-1$ degrees of freedom, a loss which cannot be made up without substantially complicating the whole procedure. Actually, if the original data consist of m sets of n observations each (n reasonably small, say, <12), the efficiency of the range estimate S_R compared to the standard pooled estimate of σ is almost 1.

3.3. *Problems Concerning the Mean.* The standard statistic for testing hypotheses concerning the mean of a normal distribution is Student's t,

$$t = \frac{\sqrt{N}(\bar{x} - \mu)}{s} \cdot$$

It seems then reasonable to investigate the distribution of a statistic of the form

$$u = \frac{\sqrt{mn}(\bar{x} - \mu)}{S_R}$$

This has been done by Lord [5] Actually, the statistic

$$G_1 = \frac{|\bar{x} - \mu|}{\bar{R}}$$

which is closely related to u requires fewer computations and is therefore preferable from a practical point of view. Critical values for this statistic have been computed by Jackson and Ross [3] from those given by Lord. If we denote these critical values at level α by $g_{1\alpha}$, the confidence interval for μ given in 2 2 follows immediately.

Again we can get an idea of the loss in accuracy due to the use of G_1 instead of t by comparing the average lengths of corresponding confidence intervals or computing a quantity similar to λ.

From 2 2, the length of the confidence interval for μ based on G_1 is $L_G = 2g_{1\alpha}\bar{R}$. Similarly, the length of the confidence interval based on t is $L_t = 2t_\alpha s/\sqrt{N}$, where t_α is the corresponding critical value for the t-distribution. Thus $E(L_G) = 2g_{1\alpha}d_n\sigma$ and $E(L_t) = 2t_\alpha c\sigma/\sqrt{N}$ where, as before, c is a constant such that $E(s) = c\sigma$.

For confidence coefficient .95, we get the following table:

N	25	50	75	100
$\frac{E(L_G)}{E(L_t)}$	1.03	1.00+	1.02	1.00+

Thus in the two cases ($N=25, 75$) when it is necessary to omit some of the original observations, $E(L_G)$ is between 2 and 3 per cent larger than $E(L_t)$, while in the other two cases the difference in length is negligible for all practical purposes [2]

From the small difference in the lengths of the two confidence intervals we may already suspect that the loss in accuracy when using the G_1-test instead of the t-test is also small. This is borne out by a study by Lord [6] Thus for $\alpha = .05$ and .01, Lord has computed quantities ρ such that when using two-sided critical regions of size α the hypothesis $\mu = \mu_0$ is rejected with probability $1-\alpha$ if actually $\mu = \mu_0 + \rho\sigma_x$ If we therefore use ρ_G and ρ_t for the G_1- and t-tests, respectively,[3]

$$\frac{\rho_G/\sqrt{mn}}{\rho_t/\sqrt{N}} = \frac{\rho_G}{\rho_t}\sqrt{\frac{N}{mn}}$$

measures by how much more the true value μ must differ from the hypothetical value μ_0 in order to be recognized by the G_1-test with the same probability $1-\alpha$ as by the t-test. For $\alpha = .05$ and the usual sample sizes, this quantity turns out to be 1 03, 1 00+, 1.02, 1.00+, respectively

Similar results hold also in the two-sample case. We may then conclude that for many practical purposes the loss in accuracy due to using G-statistics instead of the standard t-statistics is unimportant This is particularly so if the sample sizes are chosen in such a way that the computation of the mean range $\bar{R}$ does not require the omission of any observations Thus we are certainly justified in using the methods discussed in this section quite generally.

---

[2] See note 3 in this connection.

[3] In his original paper, Lord suggests that, when computing the statistic u, all observations be used in finding $\bar{x}$ in the numerator, even though it may be necessary to omit some of the observations in computing $S_{\bar{R}}$ in the denominator. If this is done, the loss in accuracy of the u-test as compared to the t-test is given by the expression ρ_G/ρ_t which for $\alpha = .05$ and the usual sample sizes take the following values 1.01, 1.00+, 1 00+ 1 00+. Thus any loss in accuracy is not so much due to the use of the mean range instead of the standard deviation, but due to the omission of some of the observations in the computation of $\bar{x}$ This loss in accuracy can be avoided if the original sample size is chosen in such a way that the subsequent analysis of the data by means of the G_1-test does not require the omission of any observations.

These remarks apply also to confidence intervals In addition, if no observations are omitted, the confidence intervals based on G_1 and t are centered at the same value $\bar{x}$, and therefore differ only in length.

REFERENCES

[1] Dixon, Wilfrid J., and Massey, Frank J., Jr., *An Introduction to Statistical Analysis.* New York: McGraw-Hill Book Company, Inc., 1951.

[2] Grubbs, Frank E., and Weaver, Chalmers L., "The best unbiased estimate of population standard deviation based on group ranges," *Journal of the American Statistical Association,* 42 (1947), 224–41.

[3] Jackson, J. Edward, and Ross, Eleanor L., "Extended tables for use with the 'G'-test for means," *Journal of the American Statistical Association,* 50 (1955), 416–33.

[4] King, E. P., "Estimating the standard deviation of a normal population," *Industrial Quality Control,* X (1953), 1–4

[5] Lord E., "The use of range instead of standard deviation in the t-test," *Biometrika,* 34 (1947), 41–67.

[6] Lord, E., "Power of the modified t-test (u-test) based on range," *Biometrika,* 37 (1950), 64–77

[7] Tippett, L. H. C., "On the extreme individuals and the range of samples taken from normal populations," *Biometrika,* 17 (1925), 364–87.

A DIAGRAMMATIC REPRESENTATION OF THE SUM OF SQUARES AND PRODUCTS

C. C. Li

University of Pittsburgh

This paper gives a geometric illustration of some of the commonest identities in the algebra of the sum of squares. The method is essentially the construction of squares of the various numbers and their means These pictures provide us with a visual representation of what the identity says. The method is later extended to represent the identities concerning the sum of products by constructing rectangles.

I. INTRODUCTION

Two of the commonest identities in algebra of the sum of squares are

$$\sum_{1}^{N} x_\alpha^2 - N\bar{x}^2 \equiv \sum_{1}^{N} (x_\alpha - \bar{x})^2, \tag{1}$$

and

$$\sum_{i=1}^{k} \sum_{\alpha=1}^{n_i} (x_{i_\alpha} - \bar{x})^2 \equiv \sum_{i=1}^{k} \sum_{\alpha=1}^{n_i} (x_{i_\alpha} - \bar{x}_i)^2 + \sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2. \tag{2}$$

The first identity refers to a single group of N numbers, x_α, $\alpha=1, 2, \cdots, N$, and $\bar{x}=\sum x/N$, the arithmetic mean of the N numbers. The second refers to the case of N numbers subdivided into k groups ($i=1, 2, \cdots, k$) with n_i numbers in the i-th group, and $n_1+n_2+\cdots+n_k=N$ Both may be demonstrated, in one way or another, by straightforward algebra. For non-mathematical beginners, however, a visual demonstration is almost always preferred. In his teaching, the writer finds it extremely useful to illustrate the relations (1) and (2) by geometric construction The usual algebraic derivation is also given in the class. The following diagrammatic representations are not meant to be formal "proofs" but are simply pictures to illustrate what the identity says

II. REPRESENTATION OF IDENTITY (1)

With the given N numbers, (i) draw a base line of length $T=x_1+x_2+\cdots x_N$, (ii) cut the base line into N segments, the first segment of length x_1, second of length x_2, etc. and (iii) construct a square for each single segment. Then the total *area* of this series of squares is $\sum x_\alpha^2=A$,

which is represented in the top row of Figure 1; in our example the numbers are $x_\alpha = 6, 2, 9, 3$, with $T = 20$, cut into 4 segments.

Next draw another base line of the same length T, but this time cut it into N segments of *equal length*. The segment length is then $\bar{x}$; in our example, $\bar{x} = 20/4 = 5$. As before, construct a square for each segment. The total area of this series of squares is $\bar{x}^2 + \bar{x}^2 + \cdots \bar{x}^2 = N\bar{x}^2 = C$, represented by the middle row of Figure 1. At this point it should be said that the area A, comprising N squares of unequal size, is always larger than the area C, comprising N squares of the same size on the same base line of length T. In other words, for any fixed length T and fixed number (N) of segments to be cut, the area $N\bar{x}^2$ is a minimum. As a matter of fact, the larger the differences between the segments,

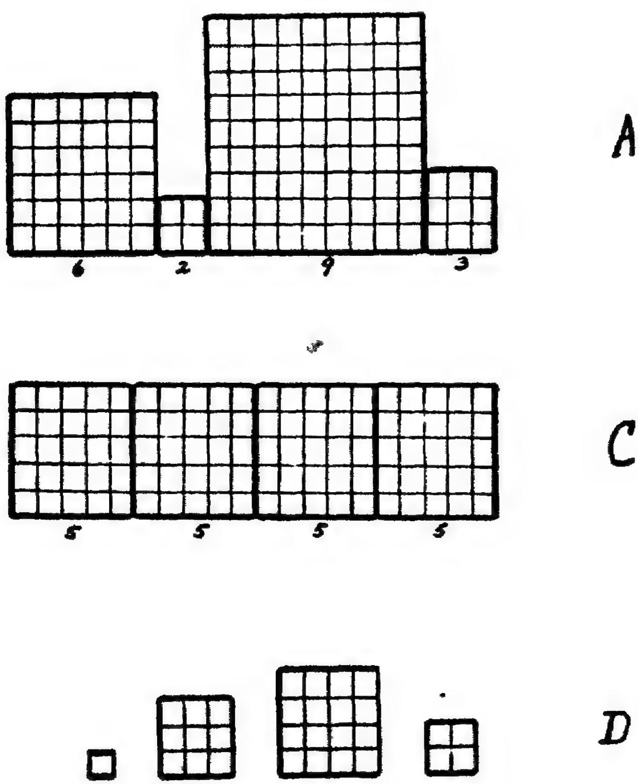


FIG. 1

the larger will be the area A. If the student is not convinced, suffice it to point out that in cutting the length 10 into two segments, $9^2+1^2>8^2+2^2>7^2+3^2>6^2+4^2>5^2+5^2$

Finally, as illustrated in the bottom row of Figure 1, take the difference $x_\alpha-\bar{x}$ for each pair of corresponding segments and construct a square for each of them. The total area of N such squares is $\sum (x_\alpha-\bar{x})^2=D$ In the example, the area is equal to

$$D = (6-5)^2 + (2-5)^2 + (9-5)^2 + (3-5)^2$$
$$= 1 + 9 + 16 + 4 = 30.$$

The area D is exactly equal to the difference between A and C. Thus, the identity (1) may be written in terms of the three areas:

$$A - C = D; \tag{1'}$$

in our example, $130-100=30$. Once this idea is clear, the last series of squares (area D) may be omitted from the diagram and simply be referred to as the difference $A-C$. This will be done in the next section

III. REPRESENTATION OF IDENTITY (2)

The chief value of the method described above lies in the ease and clarity with which the identity (2) may be demonstrated We shall take as our example the ten ($N=10$) numbers 6, 1, 2, 8, 2, 7, 6, 2, 1, 5, and shall further suppose that they are divided into three ($k=3$) groups The first group with $n_1=3$ being 6, 1, 2, the second group with $n_2=5$ being 8, 2, 7, 6, 2, and the third group with $n_3=2$ being 1, 5, so that $n_1+n_2+n_3=N=10$ Construct the areas A and C as before, regardless of the grouping; the length of the base line is $T=\sum\sum x_{i_\alpha}=40$ as shown in Figure 2. The areas are:

$$A = \sum\sum x_{i_\alpha}^2 = 224$$
$$C = N\bar{x}^2 = 160$$
$$A - C = \sum\sum (x_{i_\alpha} - \bar{x})^2 = 64$$

The last expression is the left-hand member of identity (2) and is the total sum of squares of deviations from the general mean. This value, of course, has nothing whatsoever to do with the grouping of the numbers.

Now, in between A and C, draw another base line of the same length T. Cut the base line into k "major" segments, so that the first major segment length is the total of the first group of numbers; that is, $T_1=\sum_\alpha x_{1_\alpha}=6+1+2=9$, and $T_1+T_2+\cdots+T_k=T$. Then cut each

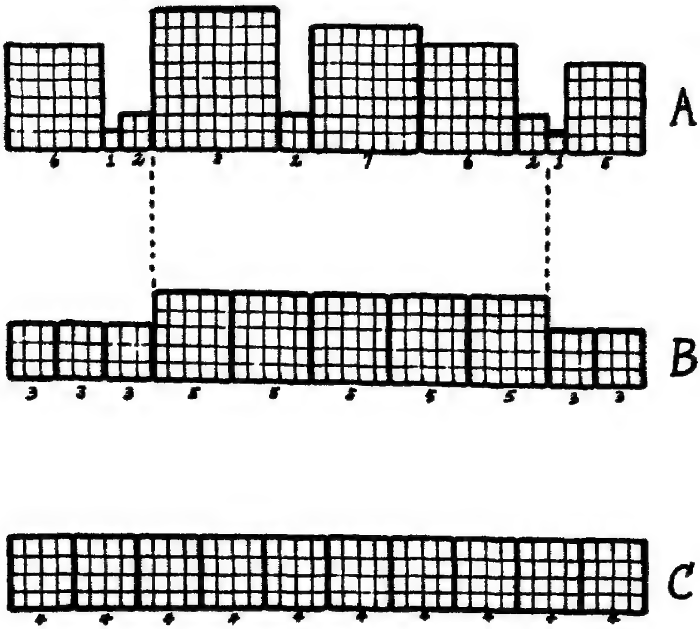


FIG. 2

major segment T_i into n_i *equal* parts, the length of these minor segments is $\bar{x}_i$, the mean value of the ith group. Construct a square on each of the minor segments The total area of these series of squares is

$$B = n_1\bar{x}_1^2 + n_2\bar{x}_2^2 + \cdots + n_k\bar{x}_k^2 = 170 \tag{3}$$

as shown in the middle row of Figure 2. By the argument advanced in Section II, we know not only that A is larger than B, but also know by how much A exceeds B. Comparing group by group, the first n_1 squares in A have an area larger than $n_1\bar{x}_1^2$ in B, and this is true for every group. Hence,

$$\begin{aligned} A - B &= \left(\sum_{\alpha} x_{1\alpha}^2 - n_1\bar{x}_1^2\right) + \cdots + \left(\sum_{\alpha} x_{k\alpha}^2 - n_k\bar{x}_k^2\right) \\ &= \sum_{\alpha} (x_{1_\alpha} - \bar{x}_1)^2 + \cdots + \sum_{\alpha} (x_{k_\alpha} - \bar{x}_k)^2. \end{aligned}$$

which is the first term of the right side of (2).

Similarly, we see that B is larger than C. Using the principle embodied in identity (1) once more, we obtain

$$B - C = n_1\bar{x}_1^2 + \cdots + n_k\bar{x}_k^2 - N\bar{x}^2 = \sum_i n_i(\bar{x}_i - \bar{x})^2.$$

In brief, any sort of grouping will yield an area intermediate between A and C, because the series of squares in B is more even in size than that in A, but not as even as that in C.

On re-examining the identity (2), we see that it says nothing more than the obvious truth that

$$A - C \equiv (A - B) + (B - C). \qquad (2')$$

In practical calculation all we need to compute are the three quantities, A, B, C.

IV. FURTHER RELATIONS

Using the general procedure of constructing squares, a number of other simple algebraic relations may be illustrated. For example, if a single big square is constrctued on the base line T of part C in Figure 1, we obtain a diagrammatic representation of the relation

$$C = N\bar{x}^2 = T^2/N.$$

Similarly, if squares are constructed for each major segment in B of Figure 2, we see that

$$B = \frac{T_1^2}{n_1} + \cdots + \frac{T_k^2}{n_k} \cdot \qquad (3')$$

The procedure of Figure 2 may be extended to the case of subgroupings within groups. The more complicated identities for the general "box-within-a-box" type of data may thus be written down immediately without going through the algebra.

V. SUM OF PRODUCTS

The diagrammatic method for the sum of squares may be readily modified to represent the corresponding identities concerning the sum of products of two series of numbers:

$$\sum xy - N\bar{x}\bar{y} \equiv \sum (x - \bar{x})(y - \bar{y}) \qquad (4)$$

$$\sum_i \sum_a (x_{ia} - \bar{x})(y_{ia} - \bar{y}) \equiv \sum_i \sum_a (x_{ia} - \bar{x}_i)(y_{ia} - \bar{y}_i) + \sum_i n_i(\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y}). \qquad (5)$$

Detailed explanation seems unnecessary. In the example we consider the following four ($N=4$) pairs of numbers:

$$x = 6, 2, 9, 3; \qquad T_x = 20,$$

$$y = 4, 1, 5, 2; \qquad T_y = 12.$$

The area of the series of *rectangles* $x \cdot y$ is (Fig. 3, top)

$$A = 24 + 2 + 45 + 6 = 77.$$

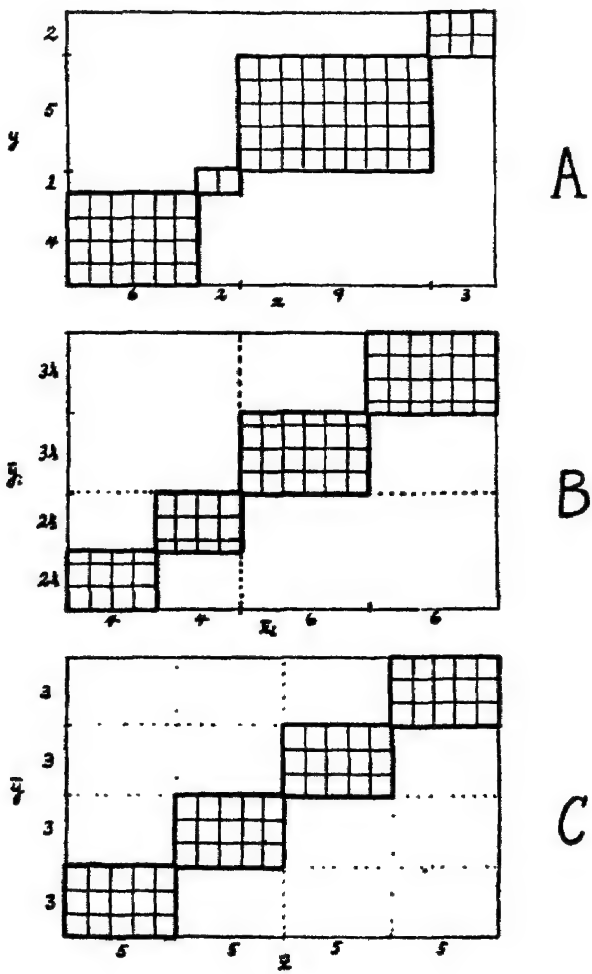


FIG 3

From the bottom diagram of Fig. 3 we see that the area of the N rectangles, each of size $\bar{x}\ \bar{y}$, is

$$C = N\bar{x}\bar{y} = \frac{T_x T_y}{N} = 60.$$

Hence,

$$\sum (x - \bar{x})(y - \bar{y}) = A - C = 17. \tag{4'}$$

Next, let us consider (5) Similar to the sum of squares, the sum of products of deviations, $A-C$, may be subdivided into two components—one for "within groups" and one for "between groups." For simplicity we divide the four pairs of (x, y) in our example into two groups, the first group comprising the first two pairs of (x, y), and the second group the last two pairs $(n_1+n_2=2+2=4)$. Using the values of $\bar{x}_i$ and $\bar{y}_i$, the group means of x and y, we construct the series of rectangles as shown in the middle diagram of Fig 3, whose area is

$$\begin{aligned} B &= n_1\bar{x}_1\bar{y}_1 \qquad + n_2\bar{x}_2\bar{y}_2 \\ &= 2\left(4 \times \frac{5}{2}\right) + 2\left(6 \times \frac{7}{2}\right) = 20 + 42 = 62. \end{aligned}$$

Again, the identity (5) simply says:

$$A - C \equiv (A - B) + (B - C). \tag{5'}$$

In our example,

$$17 = 15 + 2.$$

One important difference between the sum of squares and the sum of products is that the former is always a positive quantity while the latter can be negative The case of negative sum of products may also be visualized easily by the aid of diagrams. For instance, if the four pairs of numbers are:

$$x = 6, 2, 9, 3;$$

$$y = 2, 5, 1, 4;$$

the area of the four rectangles will be (Fig. 4)

$$A = 12 + 10 + 9 + 12 = 43,$$

which is smaller than $C = N\bar{x}\bar{y} = 60$. In this case,

$$\sum (x - \bar{x})(y - \bar{y}) = A - C = 43 - 60 = -17.$$

Any component of (5) may be positive or negative, depending on the paired values of x, y.

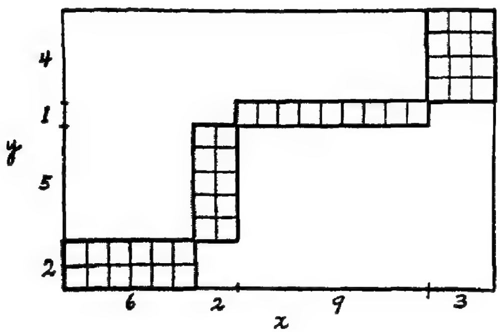


Fig. 4

THE ANALYSIS OF VARIANCE: A GRAPHICAL REPRESENTATION OF A STATISTICAL CONCEPT*

BENJAMIN A. BARNES
Massachusetts General Hospital

ELINOR PEARSON
Brooke Army Medical Center

AND

ERIC REISS
Washington University

The analysis of variance is presented in graphical form to emphasize the conventional partitioning of the sums of squares. Areas are defined representing squares of non-uniform series and sums of squares of uniform series, and the difference is displayed as the variation of the non-uniform series. The graphical presentation is extended to intra- and intergroup comparisons and to the F test. The two-dimensional analogy is presented as an aid to the teaching of statistics supplementing the algebraic formulations.

I. INTRODUCTION

The analysis of variance has been discussed in various monographs and texts on statistical analysis [1, 2, 3, 4, 5]. The mathematical formulations in an introductory presentation not uncommonly appear as somewhat of an enigma to the uninitiated. This paper illustrates a a graphical representation of a statistical concept, and the authors believe that it is a useful pedagogical technique. In no sense is it proposed that a graphical presentation should replace the mathematical arguments. However, a visual, two-dimensional analogy of certain concepts basic to the understanding of the analysis of variance may supplement the mathematical symbolism for those to whom the mathematical expressions lack a little warmth on first acquaintance.

It is well known that if a sum is divided into a given number of parts, the sum of the squares of these parts will be a minimum when all the parts are equal. From this proposition a method of measuring the variability, or lack of uniformity, in any series at once suggests itself to the student, since the difference between the sum of the squares of a non-uniform series and an appropriate uniform series might prove to be a

* The concepts presented are an extension and elaboration of material reviewed by the authors at a seminar held in July, 1952, by the Surgical Research Unit, Brooke Army Medical Center, Fort Sam Houston, Texas, under the command of Colonel William H. Ampspacher, M.D.

The authors gratefully acknowledge the helpful advice and criticism by Frederick Mosteller in the preparation of this paper.

sensitive index of the lack of uniformity. This difference of the sums of squares is, of course, a standard measure in the analysis of variance as the following well-known relationships indicate:

$$\sum(X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{N} = \sum X^2 - N\bar{X}^2.$$

In this paper the term *variation* is reserved for quantities representing the difference of the sums of squares of non-uniform and uniform series.

This orientation to the comparisons inherent in the analysis of variance is not always emphasized, and it must be clearly understood to follow the subsequent arguments. From one source [1] we present with minor modifications the formulas in the analysis together with the definitions of the symbols used in this paper:

$X_1, X_2, \cdots X_n$ and $Y_1, Y_2, \cdot Y_n$ are specific terms in Group X Data, Group Y Data, etc.;
m represents the number of groups of data,
V represents any term in any group,
$\bar{V}$ represents the mean of all terms;
$\bar{V}_K$ represents the mean of terms in the K data;
N represents the total number of terms;
N_K represents the number of terms in the K data;
$\sum$ represents a summation over the entire series,
$\sum_1^m$ indicates a summation over the m groups of data;
$\sum_1^{N_K}$ indicates a summation over the K data.

If each group has the same number of terms, then variation within groups is defined as:

$$\sum_1^m\left[\sum_1^{N_K}(V - \bar{V}_K)^2\right] = \sum V^2 - \frac{\sum_1^m\left(\sum_1^{N_K} V\right)^2}{N_K}$$
$$= \sum V^2 - \sum_1^m (N_K\bar{V}_K^2).$$

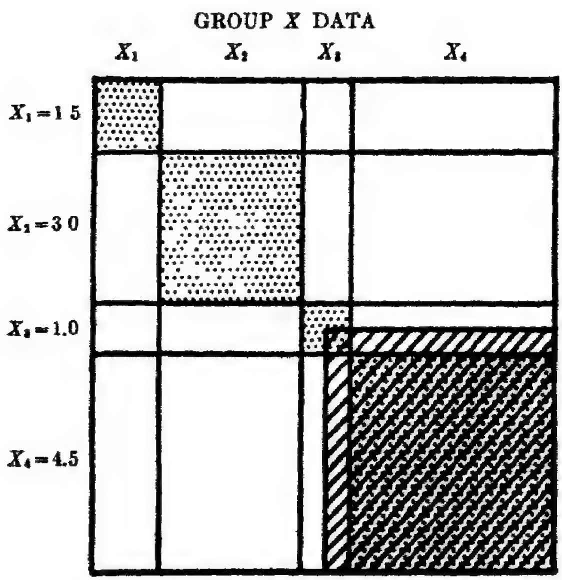
Also variation between groups is defined as:

$$\sum_1^m [N_K(\bar{V}_K - \bar{V})^2] = \frac{\sum_1^m\left(\sum_1^{N_K} V\right)^2}{N_K} - \frac{(\sum V)^2}{N}$$
$$= \sum_1^m (N_K\bar{V}_K^2) - N\bar{V}^2$$

if there are the same number of terms in the different groups. The fact that differences of squared quantities are equated with certain statistics used in the analysis of variance is clear if the formulas are expressed as above.

II. GRAPHICAL PRESENTATION

In Fig. 1 are represented Group X Data consisting of a series of 4 terms with the values as shown. These values are represented as successive distances along the edge of a square, and from the geometrical construction the squares of these values are indicated by the stippled areas running diagonally across the large square. The cross-hatched area represents the sum of the squares of a hypothetical series of *equal*



$$\bar{X} = \frac{\sum_1^4 X}{4} = \frac{10}{4} = 2.5$$

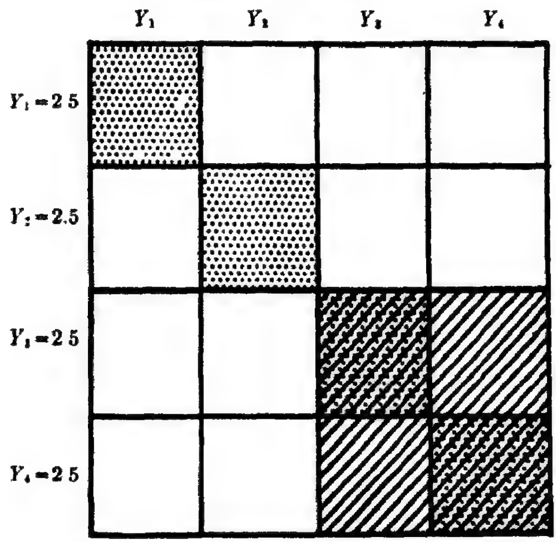
$$\sum_1^4 (X - \bar{X})^2 = \sum_1^4 X^2 - 4\bar{X}^2$$
$$= 1.5^2 + 3.0^2 + 1.0^2 + 4.5^2 - 4 \times 2.5^2$$
$$= 32.5 - 25.0 = 7.5$$

FIG. 1

numbers containing the same number of terms and having the same mean. In the expressions for the computation of variation below the figure, the first term in the difference is represented by the sum of the stippled areas and the second term is represented by the cross-hatched area, i.e., $4\bar{X}^2$. The difference of these two areas is numerically equivalent to the variation of the Group X Data. The arithmetical computation in Fig. 1 reveals this to be 7.5.

In Fig. 2 are represented Group Y Data consisting of a series of 4 terms with the values as shown. This series is peculiar in consisting of 4 identical values. By direct analogy with the discussion above, the sum of the squares of this series is equal to the area represented in the second term of the difference in the expressions below the figure. Since the difference in area is 0, the variation is likewise *nil*.

GROUP Y DATA



$$\bar{Y} = \frac{\sum_1^4 Y}{4} = \frac{10}{4} = 2.5$$

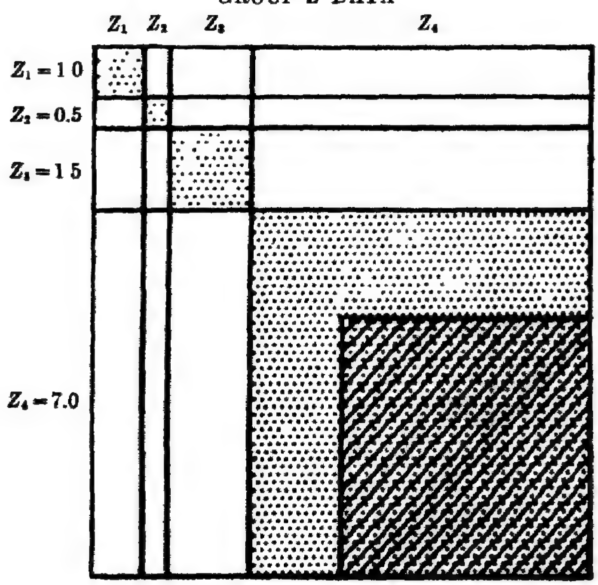
$$\begin{aligned}\sum_1^4 (Y - \bar{Y})^2 &= \sum_1^4 Y^2 - 4\bar{Y}^2 \\ &= 4 \times 2.5^2 - 4 \times 2.5^2 \\ &= 25 - 25 = 0\end{aligned}$$

FIG. 2

In Fig. 3 are represented Group Z Data consisting of a series of 4 terms of greater disparity. The variation here is 27.5.

In summary the considerations thus far have attempted to elucidate by graphical methods the implications of the difference between certain squared quantities in the analysis of variance. In each instance it is the difference between the sum of the squares of a particular series and the sum of the squares of a uniform series with an equal number of terms and sum and mean. The differences in the areas found have been displayed as the *variation* in a statistical analysis. The values of 7.5, 0, and 27.5 are each an index of the variation *within* a series. The comparison of the Group X, Y, and Z Data with each other presents a more complicated problem and a more complicated geometrical

GROUP Z DATA



$$\bar{Z} = \frac{\sum_{1}^{4} Z}{4} = \frac{10}{4} = 2.5$$

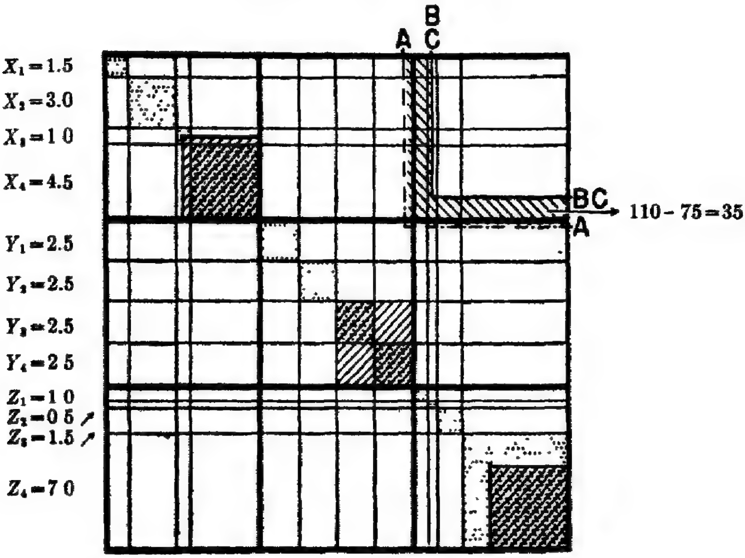
$$\begin{aligned}\sum_{1}^{4} (Z - \bar{Z})^2 &= \sum_{1}^{4} Z^2 - 4\bar{Z}^2 \\ &= 1^2 + 0.5^2 + 1.5^2 + 7.0^2 - 4 \times 2.5^2 \\ &= 52.5 - 25 = 27.5\end{aligned}$$

Fig. 3

analogy. To determine the variation *between* these three series, in distinction to the variation *within* any given series, areas must be defined representative of the sum of the squares of the terms of each series *had each series been composed of uniform values,* and the sum of these squared values must be compared to an area based on the sum of the squares of a hypothetical series of 12 equal terms with a sum and mean equivalent to the grand sum and grand mean of the combined three series.

In Fig. 4 are represented the combined data of the previous three

COMBINED GROUP DATA I



$$\bar{V} = \frac{\sum V}{N} = \frac{30}{12} = 2.5$$

Within groups: $$\sum_1^m \left[\sum_1^{N_K} (V - \bar{V}_K)^2 \right] = \sum V^2 - \sum_1^m (N_K \bar{V}_K^2)$$
$$= 110 - 3 \times 4 \times 2.5^2$$
$$= 110 - 75 = 35$$

Between groups: $$\sum_1^m \left[N_K (\bar{V}_K - \bar{V})^2 \right] = \sum_1^m (N_K \bar{V}_K^2) - N\bar{V}^2$$
$$= 3 \times 4 \times 2.5^2 - 12 \times 2.5^2$$
$$= 75 - 75 = 0$$

Fig. 4

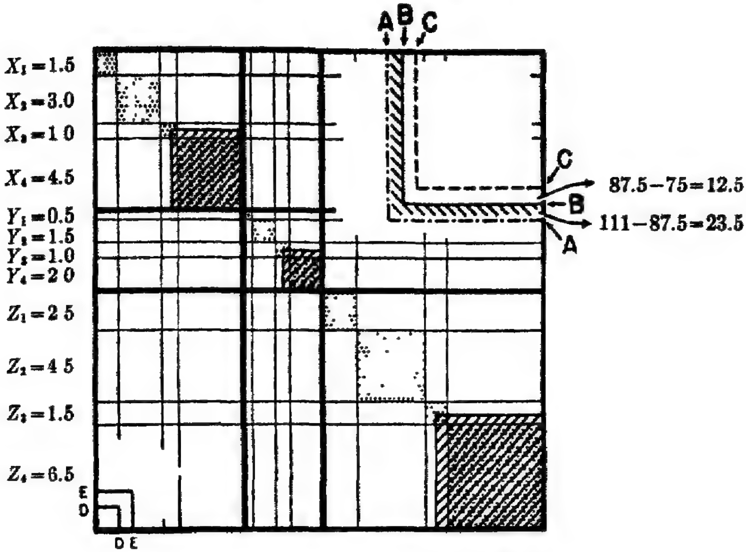
figures. In addition there are two squares in the right upper corner: the larger one designated by the dash-dotted line AA and a smaller one designated by the solid line $BCBC$. The square area between the line AA and the corner represents that obtained by adding the squares of the 12 individual terms. The square area defined by the line $BCBC$ represents the area obtained by adding the squares of 12 equal terms with a sum equal to 30. This area is also equal to one obtained from the sum of the squares of the individual terms *had each series within itself been composed of uniform values.* In other words, the area AA is equal to the sum of the stippled areas, and the area $BCBC$ is equal to the sum of the cross-hatched areas within each individual series and is also equal to the sum of the squares of 12 equal terms totalling the grand sum of 30.

Considering the formulas in Fig. 4 by inspection the *within group* (intra-group) variation is numerically equal to the area $AA\ BCBC$, an L-shaped area with a value of 35 The *between group* (inter-group) variation is numerically equal to the difference between the two equal areas represented by the square $BCBC$. Thus the between group variation is 0, and no variation exists between the three series. This is indeed confirmed by our common sense evaluation since there were four determinations in each of the three series with a common sum of 10 and a common mean of 2 5 in each series. The within group variation of 35 is equal to the sum of the variations of the three separate groups of data as previously displayed in Figs. 1–3: $7.5+0+27.5=35$.

The lack of between group variation in the above three series characterizes them as a special case, and a more generalized set of three series consisting of four terms each but with *differing* sums and means is shown in Fig. 5. The square area between the line AA and the corner is equal to the sum of the squares of the 12 individual terms in the three series and is equal to the sum of the stippled areas. The area by which this square exceeds that defined by BB, namely 23.5, is a measure of the variation *within groups*. The square defined by BB is equal to the sum of the three cross-hatched areas in each series, and the area by which this square exceeds that defined by CC, namely 12.5, is a measure of the variation *between groups*. The smallest square defined by CC is obtained by adding the squares of 12 equal terms with a sum equal to 30. The formulas and values for substitution are indicated in the figure.

In Fig. 5 the left lower corner contains two additional squares that represent the effects of the correction for the degrees of freedom upon which the variation within and between groups is based The larger square defined by EE derived from the L-shaped area $BBCC$ of 12.5

COMBINED GROUP DATA II



$$\bar{V} = \frac{\sum V}{N} = \frac{30}{12} = 2.5$$

Within groups $$\sum_{1}^{m}\left[\sum_{1}^{N_K}(V - \bar{V}_K)^2\right] = \sum V^2 - \sum_{1}^{m}(N_K \bar{V}_K^2)$$
$$= 111 - 4(2.5^2 + 1.25^2 + 3.75^2)$$
$$= 111 - 87.5 = 23.5$$

Between groups $$\sum_{1}^{m}\left[N_K(\bar{V}_K - \bar{V})^2\right] = \sum_{1}^{m}(N_K \bar{V}_K^2) - N\bar{V}^2$$
$$= 4(2.5^2 + 1.25^2 + 3.75^2) - 12 \times 2.5^2$$
$$= 87.5 - 75 = 12.5$$

F test $$\frac{\text{Square } EE}{\text{Square } DD} = \frac{12.5 \div 2}{23.5 \div 9} = \frac{6.25}{2.61} = 2.39;\ P > 0.05$$

FIG. 5

is corrected for the 2 degrees of freedom possible in the comparison of the three series and has a value of 6.25. The smaller square defined by DD derived from the L-shaped area $AABB$ of 23.5 and corrected for the 9 degrees of freedom, on which the within group comparison was based, has a value of 2.61 These areas defined by EE and DD represent the *variance* of the *between* group and *within* group compari-

sons respectively. The F test (2, 5) is the comparison of these two values, and the variance ratio with the corresponding P value is shown.

III. DISCUSSION

Since instruction in statistics is assuming a more prominent position in the curriculum of universities, pedagogical techniques must be developed to aid the student in grasping the concepts of statistical theory. Visual representations are helpful, and there is merit in stressing an interpretation of the analysis of variance which emphasizes that the sum of the squares of the parts of a whole will approach a minimum as the parts achieve increasing uniformity. From this basic proposition, a graphical representation of certain statistical concepts inherent in the analysis of variance may be developed. These techniques in no way replace the mathematical derivation of the formulas, but they may serve as a supplement.

BIBLIOGRAPHY

[1] Croxton, F. E., and Cowden, D. J., *Applied General Statistics*. New York Prentice-Hall, Inc. (1939), pp. 351–9 and 847–50.

[2] Fisher, R. A., *Statistical Methods for Research Workers*. 10th edition. Edinburgh: Oliver and Boyd (1948), Chapter 7

[3] Mainland, D., *Elementary Medical Statistics*. Philadelphia: W. B. Saunders Co. (1952), pp. 282–9.

[4] Mather, K., *Statistical Analysis in Biology*. New York: Interscience Publishers, Inc. (1946), Chapter 6.

[5] Snedecor, G. W., *Statistical Methods*, 4th edition Ames, Iowa: Iowa State College Press (1946), Chapter 10.

NOMOGRAPH FOR COMPUTING MULTIPLE CORRELATION COEFFICIENTS

Frederic M. Lord
Educational Testing Service

The accompanying nomograph[1] may be used for calculating a multiple correlation coefficient ($R_{1.23}$) from the zero-order correlations (r_{12}, r_{13}, and r_{23}).

USE OF THE CHART

If either r_{12} or r_{13} is negative, some sign changes are necessary before the nomograph can be entered. If only one of these two coefficients is negative, it must be made positive and the sign of r_{23} reversed. (This change is possible without affecting $R_{1.23}$ because it is equivalent to a simple reflection of the corresponding predictor variable.) If both r_{12} and r_{13} are negative, both must be treated as positive, the sign of r_{23} remaining unchanged.

In theory, it makes no difference which predictor variable is called variable 2 and which is called variable 3. In entering the nomograph, however, the user should always take as variable 2 the variable that has the *numerically* higher correlation with variable 1, so that $|r_{12}| \geq |r_{13}|$.

The network in the left half of the nomograph consists of a curvilinear coordinate system, similar to the ordinary rectangular Cartesian coordinate system. One variable, r_{12}, is represented by approximately horizontal curves labeled at their left extremity; the other variable, r_{13}, is represented by approximately vertical or by oblique curves, labeled at the top and intersecting the curves for r_{12}. The "curve" representing $r_{13}=0$, it may be noted, is the vertical straight line at the extreme left.

In general, all scales on the nomograph are marked off in intervals of .02. In certain areas where the curves would otherwise be too close together, however, alternate curves have been omitted, so that the curves appear at intervals of .04. Linear interpolation between the curves is all that is required. (Since $|r_{12}|$ is chosen greater than $|r_{13}|$, the user will not ordinarily wish to use the section of the graph where $|r_{12}| < .12$.)

The first step in using the nomograph is to find the point in the coordinate system at the left represented by (r_{12}, r_{13}). This is done by finding the intersection of the horizontal curve for r_{12} with the vertical or oblique curve for r_{13}. If $r_{12}=r_{13}$, the required point is to be found along the oblique straight line so labeled.

[1] Enlarged copies of the nomograph covered with laminated plastic may be obtained without charge by writing to Research Division, Educational Testing Service, 20 Nassau Street, Princeton, New Jersey.

The next step is to find the point representing the value of r_{23} on the vertical scale at the extreme right A straight-edge or ruler is then placed on the nomograph so as to pass through both this point and the previous one. It will be found important to use the lower edge of the ruler, rather than the upper, in order to avoid covering up portions of the nomograph necessary for the following step.

The network in the right half of the nomograph, like that in the left, represents a system of curvilinear coordinates. One variable, r_{23}, is represented by the approximately horizontal curves (labeled at the left and the right), the other variable, $R_{1\cdot 23}$, is represented by the vertical straight lines (labeled at top and bottom). The final step in the procedure is to start with the point just found on the vertical scale at the extreme right and follow the r_{23}-curve, passing leftwards along it until this curve again intersects the straight-edge The $R_{1\cdot 23}$-value of this intersection is the required value of the multiple-correlation coefficient.

The user will wish to cover the nomograph with some kind of waterproof plastic material on which the first point (r_{12}, r_{13}) can be marked with a dot of ink. Some method of marking this point temporarily is a virtual necessity. The ink can be readily removed from the plastic with a damp cloth (or finger)

The reader may wish to try out the procedure just described with the following examples·

r_{12}	r_{13}	r_{23}	$R_{1\cdot 23}$
.52	22	40	.520
.32	58	04	.652
.48	− 22	.18	.572
.20	02	60	.236

It will be observed in working out the last example that the value of $R_{1\cdot 23}$ cannot be determined with great accuracy from the chart because the straight-edge and the curve for r_{23} are nearly parallel at their point of intersection. This type of difficulty will be apparent to the user whenever it occurs, which is only for low values of $R_{1\cdot 23}$; it is for this reason that no lines are given for $R_{1\cdot 23} < .20$. The foregoing illustrates the fact that no general statement can be made regarding the degree of accuracy in the value of $R_{1\cdot 23}$ obtained from the chart.

ON THE CONSTRUCTION OF NOMOGRAPHS

The construction of nomographs can be a challenging although time-consuming pastime, especially if methods involving the formulation of the nomograph as a determinant are used. This method is treated in

only a minority of the texts on the subject;[2] it will be outlined very briefly here, using the present nomograph as an illustrative example, in order to indicate the general procedure.

First, the standard equation for $R_{1\cdot 23}$ is written:

$$R_{1\cdot 23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}\,. \tag{1}$$

All terms must be collected on the left side; it is convenient to square and clear of fractions before transposing, however:

$$R_{1\cdot 23}^2(1 - r_{23}^2) - r_{12}^2 - r_{13}^2 + 2r_{12}r_{13}r_{23} = 0. \tag{2}$$

Equation (2) must be written as a determinantal equation with 0 on the right side and a third-order determinant on the left. This is easily done in various ways, e.g.,

$$\begin{vmatrix} 1 & -2r_{12}r_{13} & 0 \\ r_{23} & +R^2(1 - r_{23}^2) & 1 \\ 0 & r_{12}^2 + r_{13}^2 & 1 \end{vmatrix} = 0, \tag{3}$$

as may be verified by expanding the determinant.

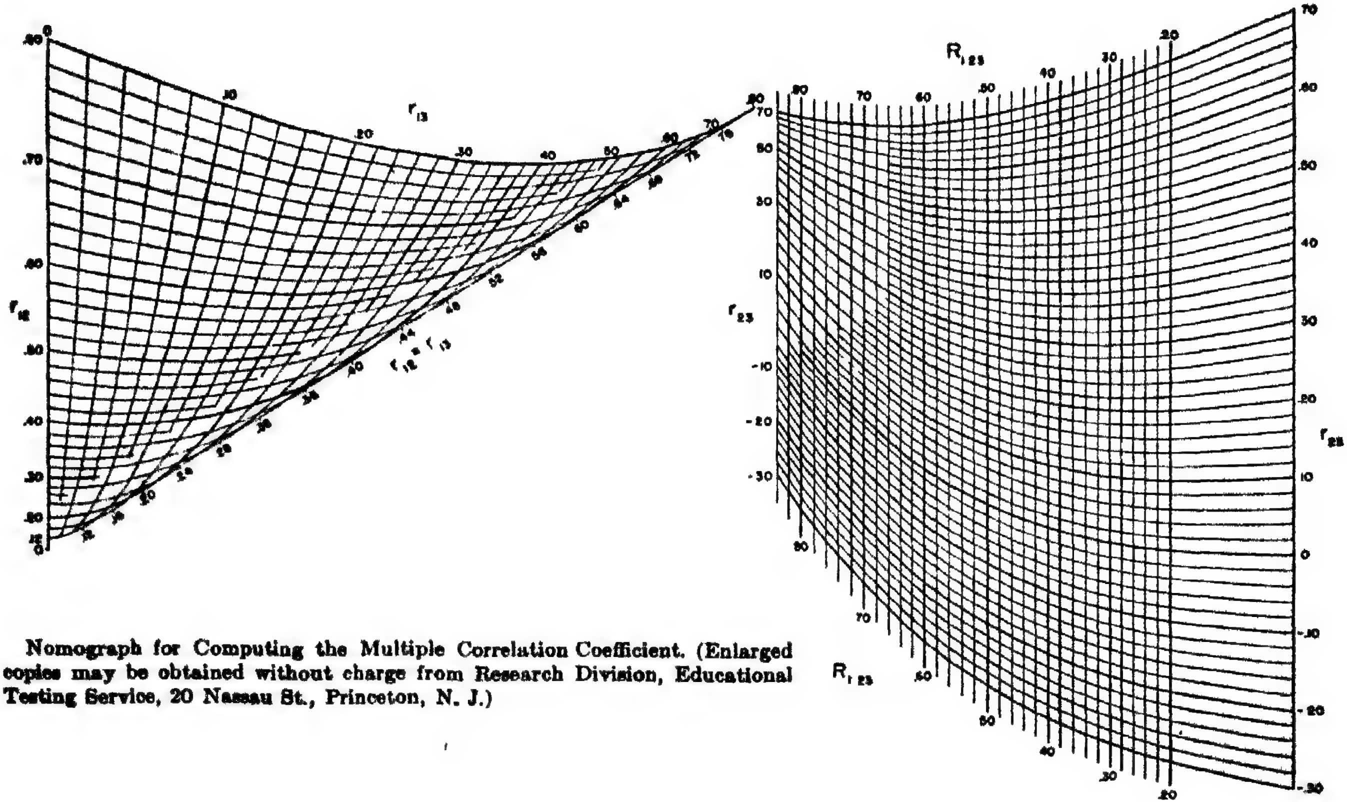
As a crucial step, equation (3) must be rewritten in the following form:

$$\begin{vmatrix} f_{11} & f_{12} & 1 \\ f_{21} & f_{22} & 1 \\ f_{31} & f_{32} & 1 \end{vmatrix} = 0, \tag{4}$$

with only two of the variables (in this case, correlation coefficients) appearing in the functions f_{ij} found in any single row. This rearrangement is achieved (if at all) by the usual procedures in which multiples of one column or row of the determinant are added to another column or row.

Finally, selected values of the variables (correlation coefficients) are chosen and the corresponding points (f_{11}, f_{12}) are plotted in ordinary Cartesian coordinates. The points determined by a single value of one variable define a curve; such a curve is drawn for each convenient value of each variable. The resulting network of curves constitutes a curvilinear coordinate system, such as appear on both the left and the right sides of the accompanying nomograph. Three such networks must be drawn corresponding to (f_{11}, f_{12}), (f_{21}, f_{22}), and (f_{31}, f_{32}). These

[2] A recent text is Allcock, H. J., and Jones, J. R., *The Nomogram* (4th Ed. Revised by J. G. L. Michel.) London: Pitman, 1950.



Nomograph for Computing the Multiple Correlation Coefficient. (Enlarged copies may be obtained without charge from Research Division, Educational Testing Service, 20 Nassau St., Princeton, N. J.)

constitute the nomograph. Ingenuity usually permits one or more of these networks to be collapsed into a single curve or straight line. In some extremely simple cases, the familiar but very specialized form of nomograph composed of three straight lines may result. In the present case, one network has degenerated into a straight-line scale, which has been made to coincide with an identical scale in one of the other networks.

If equation (1) can be expressed in the form given by equation (4) with the stated restrictions on the f's, this fact is a necessary and sufficient condition for the existence of a nomograph of the general type discussed here. If the result can be achieved at all, it can be achieved in a variety of ways, some of which are much more convenient for the user of the nomograph than others No set routine can be laid down for getting blindly from equation (3) to equation (4) without violating the restrictions on the f's, much less for obtaining an optimum result from the point of view of convenience to the user.

The determinantal equation corresponding to the accompanying nomograph is

$$\begin{vmatrix} 0 & 1.8r_{23} & 1 \\ -R_{1\cdot23} & .6\left[r_{23}(3 - 2R_{1\cdot23}) + (1 - r_{23}^2)R_{1\cdot23}^2\right] & 1 \\ -\dfrac{1.5}{r_{12}r_{13} + 1} & \dfrac{.9(r_{12}^2 + r_{13}^2)}{r_{12}r_{13} + 1} & 1 \end{vmatrix} = 0. \qquad (5)$$

APPLICATIONS OF CORRELATION MODELS FOR BISERIAL DATA

R. F. TATE*
University of Washington

This article is addressed mainly to applied workers, and is devoted to giving as complete a picture as possible of the techniques and assumptions involved when correlations are estimated from biserial data. The exposition is nontechnical in the sense that mathematical proofs are omitted and numerous examples are given. The related theoretical material is contained mainly in two recent articles by the author [11, 12]

I INTRODUCTION

THE usual model for correlation has been dealt with extensively by many authors, so a discussion of it per se will not be given. It does serve, however, as a good starting point for the present discussion, and will therefore be presented as Model A in the next section The biserial and point-biserial models will then be introduced as alterations of Model A. In sections 3 and 4 the latter models will be treated individually, together with illustrative examples Section 5 will deal with comparisons between the various models. Finally, Section 6 will close the discussion with observations on the validation of the assumptions in the biserial and point-biserial models and comments concerning which coefficient to use in a given experimental situation

II CORRELATION MODELS

Consider two continuous chance quantities X and Y, and let the parameter ρ be their population correlation. In order to speak of ρ no special assumptions need to be made about the joint probability distribution of X and Y, except that the correlation exists. However, it is a fact that for the known estimation procedures and tests of significance associated with ρ we must require that X and Y have a bivariate normal distribution. In fact, normality assumptions will be made throughout this paper, and they are all motivated by the same need, namely that of a well-defined basis for making probability statements, which in addition is easy to work with.

MODEL A

(1) Chance quantities X and Y have the bivariate normal distribution given by the bell-shaped surface

* Research sponsored by the Navy Theoretical Statistics Project, Office of Naval Research

$$\Psi(x, y) = \frac{1}{2\pi\sigma\tau\sqrt{1-\rho^2}}$$
$$\exp -\left\{\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu)^2}{\sigma^2} - \frac{2\rho(x-\mu)(y-\nu)}{\sigma\tau} + \frac{(y-\nu)^2}{\tau^2}\right]\right\},$$

where μ, ν and σ, τ are respective means and standard deviations of X and Y, and ρ is the correlation between X and Y.

(2) A random sample[1] (X_i, Y_i), $i=1, 2, \cdots, n$, is chosen from the (X, Y) population Some function of these observations is to be used to estimate ρ.

The estimate which is customarily used is $r = s_{XY}/s_X s_Y$, where

$$s_X^2 = \frac{1}{n}\sum(X_i - \overline{X})^2, \qquad s_Y^2 = \frac{1}{n}\sum(Y_i - \overline{Y})^2,$$
$$s_{XY} = \frac{1}{n}\sum(X_i - \overline{X})(Y_i - \overline{Y}).$$

The quantity r is called the (Pearson) product moment correlation coefficient

In certain situations it is impossible to observe the Y values, and yet an estimate of ρ is still desired. If the Y distribution is observable in dichotomized form, such estimation is possible. A precise description of the situation would be the following.

MODEL B

(1) Chance quantities X and Y have the bivariate normal distribution of Model A

(2) A chance quantity Z is defined by

$$Z = \begin{cases} 1 & (Y-\nu)/\tau \geqq \omega \\ 0 & (Y-\nu)/\tau < \omega, \end{cases}$$

where ω is a real number known as the point of dichotomy of the Y distribution

(3) A dichotomized random sample (X_i, Z_i), $i=1, 2, \quad , n$, is obtained from the (X, Y) population. Some function of the X and Z observations is to be used to estimate ρ

An estimate which is sometimes used is $r_b = s_{XZ}/s_X\lambda(\omega_b)$, where $\lambda(y) = 1/\sqrt{2\pi}\exp(-y^2/2)$, the normal density, and ω_b is determined by looking up in the normal table the value of ω for which

[1] Independence of observed couples is implicitly assumed in the phrase "random sample".

$$\int_{\omega}^{\infty} \lambda(y)dy = Z.$$

r_b is called the biserial correlation coefficient.

Notice that in the definition of Z the values of Y enter in the form of standard deviations from the mean. No generality will be lost, then, by choosing $\nu=0$ and $\tau=1$. We accordingly make this assumption throughout the paper. The distribution of r_b contains four unknown parameters, μ, σ, ω, ρ. In Section 4 properties of r_b and the maximum likelihood estimate $\hat{\rho}$ for Model B will be described.

Now let us go one step farther and drop the assumptions on Y altogether, dealing entirely with X and Z.

MODEL PB

(1) The chance quantity Z takes values 0 and 1, and has the distribution given by

$$g(z) = p^z(1 - p)^{1-z} \qquad z = 0, 1; \qquad 0 < p < 1.$$

(2) The chance quantity X has the *conditional* distribution[3]

$$f(x \mid z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu_z)^2/2\sigma^2}$$

When $Z=z$.

(3) A random sample (X_i, Z_i), $i=1, 2, \cdots, n$, is chosen from the (X, Z) population. Some function of these observations is to be used to estimate ρ_1, the correlation between X and Z.

Again, the customary estimate is $r_{pb}=s_{XZ}/s_Xs_Z$. r_{pb} is called the point-biserial correlation coefficient.

Considered as functions of observations, r and r_{pb}, have the same functional form. It should be remembered, however, that these two coefficients have entirely different probability distributions, since the models underlying them are quite different.

It is more convenient to discuss Model PB first, since some of the properties will be useful later when we consider Model B.

III. THE POINT-BISERIAL MODEL

It was pointed out above that r and r_{pb} have the same functional form. It follows, therefore, that certain well-known properties of r are shared by r_{pb}: namely,

[3] Note that X itself is in general not normal, but is instead a mixture of $N(\mu_1, \sigma^2)$ and $N(\mu_0, \sigma^2)$ in the proportions p and q.

(i) r_{pb} takes some value in the closed interval $[-1, +1]$ for each possible random sample (X_i, Z_i);

(ii) r_{pb} remains unchanged if the observations are coded; that is, we can write $X_i' = aX_i + b$, and r_{pb} for the random samples (X_i, Z_i) and (X_i', Z_i) will be the same;

(iii) $r_{pb} = \pm 1$ if and only if the two chance quantities are linearly related. Since Z is always 0 or 1, this means that the X observations would have to be concentrated at two points. The probability of such an occurrence is 0, but of course sample r_{pb} values of ± 1 can arise; in fact, r_{pb} is a continuous chance quantity (even though Z is discrete), so any particular value has 0 probability and all values in $[-1, +1]$ are possible as was asserted in (i). It is not possible, of course, for ρ_1 to take values ± 1. This is prohibited by the model.

With the aid of some additional notation we can express r_{pb} in a form which is probably more familiar to some readers.

$N_1 = \sum Z_i$, the number of Z observations with value 1.
$N_0 = n - \sum Z_i$, the number of Z observations with value 0.
$\hat{p} = N_1/n$
$\hat{q} = N_0/n$
$\overline{X}_1 =$ the average of the X values for the Z's which are 1.
$\overline{X}_0 =$ the average of the X values for the Z's which are 0.
$\overline{X} =$ the average of all X values.

r_{pb} can then be expressed as

$$r_{pb} = \frac{\overline{X}_1 - \overline{X}_0}{s_X} \sqrt{\hat{p}\hat{q}}$$

It is an algebraic fact that

$$ns_X^2 = \sum_{i=1}^{N_0} (X_{0i} - \overline{X}_0)^2 + \sum_{j=1}^{N_1} (X_{1j} - \overline{X}_1)^2 + n\hat{p}\hat{q}(\overline{X}_1 - \overline{X}_0)^2$$

It should be emphasized that each of the new symbols represents a chance quantity. If one thinks of the Z values as being fixed numbers, with n_1 ones and n_0 zeros, r_{pb} still has a meaning, but ρ_1 is no longer a correlation coefficient, since the possibility of correlation has been removed from the problem.

If we let the standardized difference of means $(\mu_1 - \mu_0)/\sigma$ be denoted by Δ, then ρ_1 can be written explicitly in terms of the parameters Δ and p (see [11]) as

$$\rho_1 = \Delta \sqrt{\frac{p(1-p)}{1 + p(1-p)\Delta^2}}$$

In connection with this expression for ρ_1 we should recall that Model PB specifies that the two conditional distributions of X, for $Z=0$ and $Z=1$ respectively, are normal. Δ is thus unrestricted and may take any value in $(-\infty, +\infty)$ It is then apparent that, regardless of the value of p $(0<p<1)$, ρ_1 may take any value in the (open) interval $(-1, +1)$. The (unconditional) distribution of X may itself be normal A condition which is both necessary and sufficient for X to be $N(\mu, \tau^2)$ is that $\mu_1=\mu_0=\mu$, $\sigma^2=\tau^2$. This will come up again in Section 6.

The joint probability distribution of $X_1, X_2, \cdots, X_n, Z_1, Z_2, \cdots, Z_n$ may be expressed as

$$f(x_1, \cdots, x_n; z_1, \cdots, z_n) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp - \left\{\frac{1}{2\sigma^2}\sum_{i=1}^{N_0}(x_{0i}-\mu_0)^2\right\}$$
$$\exp - \left\{\frac{1}{2\sigma^2}\sum_{j=1}^{N_1}(x_{1j}-\mu_1)^2\right\} p^{n_1}(1-p)^{n_0},$$

where $n_1=\sum z_i$, $n_0=n-n_1$, and the x_{0i} and x_{1j} represent the subdivision of x values according to the corresponding z values.

For maximum likelihood estimates, we obtain, after the usual differentiations,

$$\hat{\mu}_0 = \overline{X}_0, \quad \hat{\mu}_1 = \overline{X}_1, \quad \hat{p} = N_1/n, \quad \hat{\sigma} = \sqrt{s_X^2 - \hat{p}\hat{q}(\overline{X}_1 - \overline{X}_0)^2}.$$

Substituting these estimates in the formula for ρ_1 gives r_{pb} and thus verifies that r_{pb} is a maximum likelihood estimate of ρ_1

The exact distribution for r_{pb} is useful in some applications when small samples are considered. In general, though, for large samples the additional accuracy obtained by exact methods is not worth the extra labor. In order to use that exact distribution in the important applications it is still necessary to indulge in a small approximation. This will be made clear in connection with an example. In what follows we will describe the "exact" method, illustrate and then consider large sample approximation procedures

Let us consider the statistic[3]

$$T = r_{pb}\sqrt{(n-2)/(1-r_{pb}^2)}.$$

If we knew the distribution of T, we could if necessary find the distribution of r_{pb} from it. It is actually unnecessary for applications to know the distribution of r_{pb}; that of T will suffice. Moreover, the distribution of T is not difficult to arrive at. It turns out to be a combination (mixture) of non-central t distributions.

[3] This transformation was introduced by Lev [6].

A non-central t distribution[4] contains two parameters, the number of degrees of freedom d and the parameter of non-centrality δ. In our situation there will be a parameter of non-centrality $\delta(n_1)$ for each possible value of n_1 (see [11]):

Let $\zeta(t, \delta(n_1))$ be a non-central t distribution with $d=n-2$ and

$$\delta(n_1) = \Delta\sqrt{\frac{n_1(n-n_1)}{n}}\,.$$

Then T has the distribution

$$\xi(t) = \sum_{n_1=0}^{n}\binom{n}{n_1}p^{n_1}(1-p)^{n-n_1}\zeta(t, \delta(n_1))$$

Note that $\xi(t)$ contains the unknown parameters p and Δ, and that if it is desired Δ may be reexpressed in terms of p and ρ_1.

An important type of problem which may be handled with "exactness" by the distribution of T is that of determining whether r_{pb} is significantly different from 0. In other words, we can test the hypothesis $H:\rho_1=0$ When hypothesis H is true, then $\rho_1=\Delta=0$ and $\xi(t)$ becomes the ordinary t distribution with $n-2$ degrees of freedom (see [11] and Lev [6]). If the level of significance α is fixed, then the two-sample t test is reject H when $|T|\geqq k_\alpha$, where k_α is obtained from the t table The power function of this test, using level of significance α, is

$$\beta(p, \rho_1) = 1 - \sum_{n_1=0}^{n}\binom{n}{n_1}p^{n_1}(1-p)^{n-n_1}\int_{-k_\alpha}^{+k_\alpha}\zeta(t, \delta(n_1))dt$$

Now, if p is specified and ρ_1 is specified, we can use the table of Neyman and Tokarska [8] (provided $\alpha=.02$ or 10) to calculate $\beta(p, \rho_1)$, the probability that the test would reject the null hypothesis under these conditions. We must express $\delta(n_1)$ in the form of

$$\rho_1\sqrt{n_1(n-n_1)/np(1-p)(1-\rho_1^2)}.$$

Example 1: Let X be the total income of a household; let Z be 1 if at least one member of the household attends the theater during a given week and 0 otherwise. Suppose that 10 households are surveyed with the result that $r_{pb}=+.6624$ At the .10 level of significance do we conclude there is a connection between total income and theater entertainment? What is the probability of making the right decision if in reality $p=\frac{1}{2}$ and $\rho_1=2/\sqrt{2}=+.707$?

[4] Johnson and Welch [5] have discussed this distribution at length with tables and examples.

$$d = n - 2 = 8, \qquad \alpha = .10, \qquad k_a = 1.860$$

$$T = r_{pb}\sqrt{(n-2)/(1-r_{pb}^2)} = 2.501.$$

The answer to the first question is therefore yes. To answer the second question we construct Table I, filling in column (4) from Table I of Neyman and Tokarska [8]. We are considering a two-sample t-test with level of significance $\alpha = .10$. Their table is constructed for a one-sided t-test with significance level $\alpha = .05$. It is current practice to use their table, although small errors will creep in, due to the fact that the non-central t distribution is skewed. The error of any entry in column (4) will be smaller for smaller entries (see p. 326 [8]). It can be seen from our Table I that large entries in column (4) will in general be multiplied by small entries in column (2) in the sort of problem we are discussing Therefore, if p is not close to 0 or 1, we should expect good accuracy. Also, the first entry, .900, is exact (the Neyman-Tokarska table would have yielded a slightly higher value), so no error has arisen here. The probability of making a type II error when $p = \frac{1}{3}$, $\rho_1 = 1/\sqrt{2}$ is the sum of the last column. Therefore, the answer to the second question is

$$\beta\left(\frac{1}{3}, \frac{1}{\sqrt{2}}\right) = 1 - P(\text{Type II error}) = .832.$$

For large samples the following approximation holds (see [11]): r_{pb} *is an approximately normally distributed chance quantity with mean* ρ_1

TABLE 1

n_1	$\binom{10}{n_1}_{(1/3)^{n_1}(2/3)^{10-n_1}}$	$\delta(n_1) = \sqrt{.45n_1(10-n_1)}$	$\int_{-k_a}^{+k_a} f(t, \delta(n_1))dt$	Col (2) × Col 4
0	.0173	0	.900	.016
1	.0867	2.012	.425	.037
2	.1951	2 683	.214	.042
3	.2601	3.074	.129	.034
4	.2276	3.286	.090	.020
5	.1366	3.354	.083	.011
6	.0569	3.286	.090	.005
7	.0163	3.074	.129	.002
8	.0030	2.683	.214	.001
9	.0003	2.012	.425	.000
10	.000	0	.950	.000

and standard deviation

$$\sigma_{r_{pb}} = (1 - \rho_1^2)\sqrt{\frac{\rho_1^2 + 2p(1-p)(2 - 3\rho_1^2)}{4np(1-p)}}\,.$$

When values for p and ρ_1 are unknown and we want $\sigma_{r_{pb}}$, we have no alternative but to use the approximate values $\hat{p}$ and r_{pb}.

Example 2: Let X be the IQ of an individual. Let Z be 1 if the individual votes at a given presidential election and 0 if he does not. Assume that the people under consideration are citizens with a permanent residence in some state. Consider a sample of 100 individuals in which 65 voted and 35 did not vote, and for which $r_{pb} = +.40$. Test the hypothesis that $\rho_1 = \frac{1}{4}$ at significance level .01, and find 99% confidence limits for ρ_1.

$$n = 100, \qquad n_1 = 65, \qquad n_0 = 35, \qquad \alpha = .01$$

$$\sigma_{r_{pb}} \cong \frac{.84}{20}\sqrt{\frac{16 + .455(1.52)}{2275}} = .0813$$

From the normal distribution we obtain the figure 2.58, so 99% confidence limits for ρ_1 are $.40 \pm 2.58\sigma_{r_{pb}}$ or $(+.19, +.61)$. To test the hypothesis $\rho_1 = \frac{1}{4}$ at the .01 level we observe that .25 lies in the confidence interval, so we accept.

If the value of p is known, a definite improvement can be made on the above procedure, even for samples only moderately large. This observation stems from the fact that if p is known, the standard deviation contains only the parameter ρ_1. By a technique first used by R. A. Fisher ([3], pp. 197–201), and known as a "variance stabilizing transformation," we can transform r_{pb} and obtain a new statistic whose large-sample standard deviation depends only on the sample size. Such transformations will exist for most fixed values of p. We give the transformation for the case $p = \frac{1}{2}$ for the following reasons. First, situations in which the characteristic measured by Z is sex are among those which most clearly can be treated with Model PB, and p would in such cases represent the proportion of males or females in the population. Secondly, the large-sample standard deviation $\sigma_{r_{pb}}'$ takes on its smallest possible value (when ρ_1 is fixed) at $p = \frac{1}{2}$. For this case the transformed statistic is given by[5]

$$\phi(r_{pb}) = \operatorname{sgn}(r_{pb})\,\frac{1}{2}\log_e \frac{1 + \sqrt{1 - (1 - r_{pb}^2)^2}}{1 - \sqrt{1 - (1 - r_{pb}^2)^2}}\,.$$

[5] sgn (r_{pb}) is -1 for $r_{pb} < 0$, 0 for $r_{pb} = 0$, and $+1$ for $r_{pb} > 0$.

The result we use, then, is that *the chance quantity $\phi(r_{pb})$ is approximately normally distributed with mean $\phi(\rho_1)$ and standard deviation $\sqrt{2/n}$ when $p=\frac{1}{2}$ and n is moderately large, say* 25.

Example 3: Suppose two independent samples, containing 25 and 30 pairs of observations respectively, result in values $r_{pb}^{(1)}=+.20$ and $r_{pb}^{(2)}=-.10$ Assume $p^{(1)}=p^{(2)}=\frac{1}{2}$ for the two populations. Is there ample evidence at significance level $\alpha=.05$ to accept the hypothesis that the two correlation coefficients $\rho_1^{(1)}$ and $\rho_1^{(2)}$ are equal? Using the symbols F_1 and F_2 for the two required values of Fisher's z, we have that

$$\operatorname{sgn}(r_{pb}^{(1)})F_1 = \operatorname{sgn}(r_{pb}^{(2)})F_2$$

is approximately normal with mean 0 and standard deviation $\sqrt{2/25+2/30}=.383$ when the hypothesis is true.

From Table VIII of Pearson [9], we obtain

$$\sqrt{1-(1-.20^2)^2}=\sqrt{.0784}, \qquad \sqrt{1-(1-.10^2)^2}=\sqrt{.0199}.$$

The square roots are respectively .280 and .141. Using Table VB of Fisher [3], we have

$$\operatorname{sgn}(+.20)(.288)-\operatorname{sgn}(-.10)(.142)=+.430$$

From the normal table we obtain 1.96 for $\alpha=.05$. $.430<1.96\ (.383)$, so we accept the hypothesis.

Example 4. For an example illustrating the use of the transformed statistic $\phi(r_{pb})$ for calculating confidence limits for ρ_1 see [11] The data there are for the situation $X=\text{IQ}$ and $Z=1$ for male, 0 for female

IV THE BISERIAL MODEL

Since Pearson used the "method of consistency" to develop r_b, it is certainly a consistent estimate of ρ; that is, the probability that r_b will deviate from ρ by any preassigned amount becomes small as the sample size increases

In connection with Model PB it was noted that r for Model A and r_{pb} for Model PB had the same functional form. They were used with different assumptions on the sample, and so possessed different probability distributions. In connection with Model B we will need to speak of the sample product moment correlation coefficient for the sample (X_i, Z_i). In order to be consistent with the above convention we introduce for this statistic a new symbol, r'. Corresponding to r' we have the population parameter $\rho'=\text{Corr}\,(X, Z)$. r' and ρ' should not be con-

fused with r_{pb} and ρ_1. Now, r_b has an alternate expression, similar to that for r_{pb} and using r': namely,

$$r_b = \frac{\overline{X}_1 - \overline{X}_0}{s_X} \frac{\hat{p}\hat{q}}{\lambda(\omega_b)} = r' \frac{s_Z}{\lambda(\omega_b)}.$$

It has been known for some time that r_b may exceed 1 in magnitude for some samples, although this alone does not prevent the statistic from being used, because such an event rarely occurs There follows a short discussion of the magnitude of r_b.

First, there is the question of what happens in the limit[6] as $n \to \infty$. Population parameters ρ and ρ' are related as

$$\rho = \rho' \frac{\sqrt{p(1-p)}}{\lambda(\omega)}$$

It can be shown (see [12]) that both $\sqrt{p(1-p)}/\lambda(\omega)$ and its sample counterpart $s_Z/\lambda(\omega_b)$ are never less than $\sqrt{\pi/2}$ Thus, since $-1<\rho<+1$,

$$-\sqrt{2/\pi} < \rho' < +\sqrt{2/\pi}.$$

However, in the case of r_b and r' we see that for any fixed sample size $-1 \leqq r' \leqq +1$ (the reasoning is the same as for the case of r_{pb}). Thus,

$$\begin{aligned} |r_b| &\leqq 1 \quad \text{when} \quad |r'| \leqq \sqrt{2/\pi} \\ r_b &< -1 \quad \text{when} \quad r' < -\sqrt{2/\pi} \\ r_b &> +1 \quad \text{when} \quad r' > +\sqrt{2/\pi}. \end{aligned}$$

There is, then, an essential difference between the limiting case and the fixed sample case The fact worth noting is that not only do we have, for certain sample sizes, samples for which $r_b > 1$, but the probabilities will be positive: In order to clear up this point the following illustration is offered

Example 5: Select arbitrarily a number greater than 1, say 4 Let $n=1000$ and p, ω, ρ be subject only to their defining assumptions, $0<p<1$, $-\infty<\omega<+\infty$, and $-1<\rho<+1$. The probability that the chance quantity $\hat{p}$ takes the value 999 is $1000\, p^{999}q$. This $\hat{p}$ value yields $s_Z/\lambda(\omega_b) = 9.6$ At the same time it can be seen, subject to the condition $\hat{p} = .999$, that we have $P(r' > 417 \mid \hat{p} = .999)$ positive. Multiplication of .417 and 9.6 shows that $P(r_b > 4)$ is positive.

[6] What is referred to here is the stochastic limit, or limit in probability. The point is that r_b and r' are consistent estimates for ρ and ρ' respectively

It is obvious that the result of Example 5 holds for any (fixed) n greater than 1000, although the positive probability, $P(r_b>4)$, decreases with increasing n. This is consistent with the fact that in the limit ρ must be <1 according to the model. The reason a large n is used for the example is that we need a positive probability for $\hat{p}$ to be near 1 (and yet not equal to 1!).

Soper [10] derived the large-sample standard deviation for r_b. His expression is

$$\sigma_{r_b}=\frac{1}{\sqrt{n}}\sqrt{\rho^4+\rho^2\left[\frac{p(\omega)q(\omega)\omega^2}{\lambda^2(\omega)}+\frac{(2p(\omega)-1)\omega}{\lambda(\omega)}-\frac{5}{2}\right]+\frac{p(\omega)q(\omega)}{\lambda^2(\omega)}},$$

where as before $\lambda(\omega)=1/\sqrt{2\pi}\exp(-\omega^2/2)$, $p(\omega)=P(Y\geqq\omega)=P(Z=1)$, and $q(\omega)=1-p(\omega)$ For a table of $\sigma_{r_b}\sqrt{n}$ see Table I of [12].

It may be shown (see [12], Theorem X) that σ_{r_b} takes on its smallest value (for ρ fixed) at $p(\omega)=\frac{1}{2}$, that is when $\omega=0$. This may also be confirmed by a glance at the table just mentioned above.

As the reader may have supposed, it may be shown (see [12], Theorem VI) that for large samples r_b *is an approximately normally distributed chance quantity with mean* ρ *and standard deviation given by* σ_{r_b}

Again we have a situation with too many parameters, namely ω and ρ. If we make the substitutions $\rho=r_b$ and $\omega=\omega_b$ in σ_{r_b}, approximate answers may be obtained as in the case of r_{pb}, Example 2.

Example 6: Let X be the score of an individual on a comprehensive test in mathematics which is given at the end of 4 years of college. Let Z be 1 if a certain question on a freshman mathematical aptitude test is answered correctly and 0 if it is not. Suppose that a sample of 200 students gives the results $n_1=84$, $n_0=116$, $r_b=-.20$. Using level of significance $\alpha=.01$, do we have reason to believe that this particular test item should be omitted from the battery of test questions?

The situation described calls for a test of $H:\rho=0$. Using the data $n=200$, $\hat{p}=.42$, $r_b=-.20$ and Table I of [12], we obtain

$$\sigma_{r_b}=\frac{1.224}{\sqrt{200}}=.0866.$$

$|-.20|<2.58\,\sigma_{r_b}=.222$, so we accept hypothesis H and assert that the item should not be used.

Due to the minimal property of σ_{r_b} at $\omega=0$ our attention is again focused on this case. For Model B we obtain the variance stabilizing transformation

$$\psi(r_b) = \frac{1}{2}\log_e \frac{1 + (2r_b/\sqrt{5})}{1 - (2r_b/\sqrt{5})}$$

and it can be said that $\psi(r_b)$ *is approximately normally distributed with mean* $\psi(\rho)$ *and standard deviation* $\sqrt{5/4n}$.

Example 8: Using the data and conditions of Example 7 we will use the transformation $\psi(r_b)$ together with the assumption that $p = \frac{1}{2}$.

From Table VB of Fisher [3] we obtain

$$\frac{1}{2}\log_e \frac{1 + (-.1789)}{1 - (-.1789)} = -\frac{1}{2}\log_e \frac{1 + .1789}{1 - 1789} = -.181.$$

$|-.181| < 2.58\sqrt{5/4n} = .204$, and this time we have the same conclusion, accept $H: \rho = 0$.

The case of r_b yielding an impossible value of ρ could be taken care of, after a fashion, by truncating r_b at ± 1 A much more serious defect is that r_b lacks satisfactory large-sample efficiency. That is, its large-sample standard deviation is, for certain values of ρ, much larger than $\sigma_{\hat{\rho}}$ the large-sample standard deviation of the maximum likelihood estimate $\hat{\rho}$. In order to express results more precisely let us denote the efficiency of r_b by

$$e(\omega, \rho) = \frac{\sigma_{\hat{\rho}}^2}{\sigma_{r_b}^2} \cdot$$

The following facts concerning $e(\omega, \rho)$ may be demonstrated (see [12], Theorems VII, VIII):

$$(1) \qquad e(\omega, 0) = 1$$

$$(2) \qquad e(\omega, -1) = e(\omega, +1) = 0.$$

The facts are, then, that from the point of view of large-sample efficiency, r_b is about as good as $\hat{\rho}$ if ρ happens to be near 0, but if ρ is near ± 1, $\hat{\rho}$ is definitely preferable as an estimate of ρ [7] Since we generally don't have advance information on ρ, we should use $\hat{\rho}$ whenever time and money permit us to compute it.

The above considerations bring us to the question of how to compute ρ. This is conceptually easy, but in practice requires a fair amount of calculation. No closed form exists for the maximum likelihood estimates $\hat{\omega}$ and $\hat{\rho}$, so we resort to an iterative procedure The method to be used is Newton's method in two variables. The likelihood equations

[7] $\hat{\rho}$ is thus much superior to r_b in item evaluation in those cases in which large values of $|\rho|$ appear

in μ, σ, ω, ρ can be reduced (see [12], Section 6) to

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma} = s_x, \quad \sum (2z_i - 1)\phi_i = 0, \quad \sum x_i'(2z_i - 1)\phi_i = 0,$$

where

$$x_i' = \frac{x_i - \bar{x}}{s_x}, \qquad \phi_i = \phi\left(\frac{(2z_i - 1)(\omega - \hat{\rho}x_i')}{\sqrt{1 - \hat{\rho}^2}}\right),$$

with ϕ defined by $\phi(x) = \lambda(x)/p(x)$. In this connection recall from Section 2 that $p(x) = \int_x^\infty \lambda(y)dy$.

ITERATIVE METHOD

(1) For initial guesses of $\hat{\omega}$ and $\hat{\rho}$ use ω_b and r_b (unless $|r_b| > 1$, in which case see Remark 1 below)

(2) Compute the following for each $i = 1, 2, \cdots, n$.

$$x_i', \qquad \delta_i = 2z_i - 1, \qquad \gamma_i = \frac{\omega_b - r_b x_i'}{\sqrt{1 - r_b^2}}, \qquad \phi_i = \phi(\delta_i\gamma_i),$$

$$\phi_i - \delta_i\gamma_i, \qquad \delta_i x_i'\phi_i, \qquad A_i = \phi_i(\phi_i - \delta_i\gamma_i),$$

$$A_i x_i', \qquad A_i x_i'^2,$$

Values of ϕ_i may be obtained from the table of Birnbaum and Hartley [2].

(3) Evaluate the three determinants

$$D_1 = \begin{vmatrix} \sum A_i & r_b\omega_b \sum A_i - \sum A_i x_i' \\ \sum A_i x_i' & r_b\omega_b \sum A_i x_i' - \sum A_i x_i'^2 \end{vmatrix}$$

$$D_2 = \begin{vmatrix} -\sum \delta_i\phi_i & r_b\omega_b \sum A_i - \sum A_i x_i' \\ -\sum \delta_i x_i'\phi_i & r_b\omega_b \sum A_i x_i' - \sum A_i x_i'^2 \end{vmatrix}$$

$$D_3 = \begin{vmatrix} \sum A_i & -\sum \delta_i\phi_i \\ \sum A_i x_i' & -\sum \delta_i x_i'\phi_i \end{vmatrix}.$$

(4) The first stage will result in values

$$\omega_2 = \omega_b + \frac{D_2}{D_1}\sqrt{1 - r_b^2}$$

$$\rho_2 = r_b + \frac{D_3}{D_1}(1 - r_b^2)^{3/2}$$

Repeat the whole process, if necessary, using ω_2, ρ_2 in place of ω_b, r_b.

Remarks:

(1) If $r_b < -1$ or $r_b > +1$, we can use $-.90$ or $+.90$ respectively as a first guess for $\hat{\rho}$, together with ω_b as a first guess for $\hat{\omega}$. Other possibilities for $\hat{\rho}$, which lead to convenient calculations, are $\sqrt{.91}$ and $\sqrt{.96}$.

(2) When r_b is near ± 1, the calculations will of course be more difficult. In this case 2 or 3 stages of the procedure may be necessary. When r_b is moderate in size, 1 stage will often be sufficient.

Example 9: An example using 20 observations taken from a bivariate normal population with $\rho = 1/\sqrt{2}$, and resulting in $r_b = +.410$, is given in Section 6 of [12]. One stage is required to produce the answer $\hat{\rho} = +.489$ which is correct in the third decimal.

Example 10: As an example of a difficult case consider the following data (Again we may think of X as IQ with Z denoting pass or fail):

X	90	96	100	107	110
Z	0	0	1	1	1

Initial calculations yield $\omega_b = -.256$, $\bar{x} = 100.6$, $s_x = 7.26$, $r' = .8553$. Recalling that $r_b \geqq r'\sqrt{\pi/2}$, we see $r_b \geqq 1.072$. Therefore, assuming $r_b = +.90$ as an initial guess, together with $\omega_b = -.256$ we could start the procedure. It turns out in this case that the resulting values of stage 1 are $\omega_1 = -.3814$ and $\rho_1 = +1.025$. This indicates that a larger guess on $\hat{\rho}$ is called for. Accordingly, choose $\rho = \sqrt{.96}$ as a next guess and $\omega = -.3814$. The following are the calculations for stage 2:

x_i'	δ_i	γ_i	ϕ_i	$\delta_i x_i' \phi_i$	$\phi_i - \delta_i \gamma_i$	A_i
−1.460	−1	+5.246	.000001	+.000001	+5.246001	.000005
−0.633	−1	+1.194	.221315	+.140092	+1.415315	.313230
−0.083	+1	−1.500	.138790	−.011520	+1.638790	.227448
+0.881	+1	−6.223	0	0	+6.223000	0
+1.294	+1	−8.246	0	0	+8.246000	0
				+.128573		+.540683

$A_i x_i'$	$A_i x_i'^2$
−.000007	.000010
−.198275	.125508
−.018878	.001567
0	0
0	0
−.217160	+.127085

$$\sum \delta_i \phi_i = -.082526 \qquad D_1 = -.021068$$

$$\rho_2\omega_2 \sum A_i - \sum A_i x_i' = +.015109 \qquad D_2 = -.001773$$

$$\rho_2\omega_2 \sum A_i x_i' - \sum A_i x_i'^2 = -.045033 \qquad D_3 = -.051596$$

$$\omega_2 = \omega + \frac{D_2}{D_1}\sqrt{1-\rho^2} = -.3646$$

$$\rho_2 = \rho + \frac{D_2}{D_1}(1-\rho^2)^{3/2} = +.9994.$$

The true value of ρ therefore lies between $\sqrt{.96} = +.9798$ and $+.9994$.

Analytic expressions for $\sigma_{\hat{\omega}}^2$ and $\sigma_{\hat{\rho}}^2$ are available (see [12], Theorem I), but have not been tabulated.

As for r_b and r_{pb}, we again find that $\hat{\rho}$ is approximately normal for large samples and that $\sigma_{\hat{\rho}}^2$ is smallest when $\omega = 0$.

V. COMPARISON BETWEEN MODELS

Setting aside for the moment any questions we may have concerning the validity of the Models A, B, and PB in given experimental situations, let us make all of the appropriate comparisons we can.

(1) Since all coefficients are estimates of population correlations, they are measures of the degree of linear relationship, and not necessarily of general dependence.

(2) Under the conditions of Models B and Model PB r_b, $\hat{\rho}$ and r_{pb} are all unchanged by coding of the X values.

(3) r, $\hat{\rho}$, and r_{pb} are maximum likelihood estimates for ρ, ρ, and ρ_1 respectively in the case of Models A, B, and PB. In addition r_b is essentially as good as a maximum likelihood estimate when $\rho \cong 0$.

(4) For large samples $\sigma_{r_b} \geqq \sigma_{\hat{\rho}} > \sigma_r$. The first inequality holds because $\hat{\rho}$ is a maximum likelihood estimate. The second inequality holds because Model B gives restricted information for the Y observations of Model A and r is maximum likelihood for Model A.

(5) Nothing is gained by comparing σ_{r_b} or $\sigma_{\hat{\rho}}$ with $\sigma_{r_{pb}}$, since r_b and $\hat{\rho}$ estimate one parameter and r_{pb} estimates another.

(6) r, r_{pb}, and $\hat{\rho}$ are all restricted to the interval $[-1, +1]$, whereas r_b is unbounded.

(7) σ_{r_b}, $\sigma_{\hat{\rho}}$, and $\sigma_{r_{pb}}$ all take their smallest values (considering ρ and ρ_1 fixed) when $p = \frac{1}{2}$, that is when the Z observations take the value 1 on the average 50% of the time.

(8) A variance stabilizing transformation is available for r, r_b, r_{pb}, and Table VB of Fisher [3] can be used in all cases.

VI. ASSUMPTIONS UNDERLYING MODELS B AND PB

There is considerable controversy over the question of when to use "biserial r" and when to use "point-biserial r." In terms of this paper the question becomes that of when to use Model B and when to use Model PB.

Assumptions concerning the X observations in Models B and PB may be examined by using known statistical tests. The matter of deciding whether or not the Z observations may be considered to stem from a dichotomy of a continuous distribution is, on the other hand, a fairly open question which depends for the most part on what is known about the data at hand, or what may be assumed by invoking theoretical concepts of the particular field of application. Comments will be made below in connection with this problem. It is hoped, at least, that certain misconceptions may be removed from consideration.

First consider Model PB. This is the most satisfying from the point of view of a mathematical statistician. The quantity r_{pb} is a maximum likelihood estimate; it is easy to compute. Moreover, in Section 3 it was seen that clear cut estimation procedures exist, especially for the case $p=\frac{1}{2}$. In any situation which produces biserial data, we can use ordinary tests of normality to investigate the assumptions on the conditional distributions which arise in Model PB From Section 2 the probability distribution of an X observation is

$$h(x) = p \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma^2} + (1-p) \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu_0)^2/2\sigma^2}.$$

One method of testing goodness of fit to the distribution $h(x)$ is to use normal probability paper [8] This is done as follows.

(1) Compute $\bar{x}_0$, $\bar{x}_1$, s

(2) Using probability paper plot 2 curves. For observations x_{01}, x_{02}, $\cdots$, x_{0n_0} plot $1/n_0$ (Number of observations $\leqq x_{0i}$) against $(x_{0i}-\bar{x}_0)/s$. For observations x_{11}, x_{12}, $\cdots$, x_{1n_1} plot $1/n_1$ (Number of observations $\leqq x_{1i}$) against $(x_{1i}-\bar{x}_1)/s$

(3) If both curves are approximately straight lines, conclude that Model PB is satisfactory.

The above tells us only whether the normality conditions for Model PB are satisfied. It doesn't indicate anything about whether or not ρ_1 is the quantity we should be interested in.

In the case of Model B we compute $\bar{x}$, s_x and apply the considerations of the previous paragraphs. There is, of course, the possibility that the conditions of both Model B and Model PB will be satisfied.[9] In such a case we would have $\rho=\rho_1=0$. ρ_1 would be 0 since normality of X in

[8] It has been shown in a recent article by Chernoff and Lehmann [1] that the chi-square statistic does not have the χ^2 distribution for large samples, as was previously thought, if maximum likelihood estimates based on the full sample are used. In our problem the state of affairs is that if we pick a significance level $\alpha=.05$ and use chi-square we wrongly reject the distribution $h(x)$ more than 5% of the time.

[9] Notice that in this case ρ_1 of Model PB and ρ' of Model B (introduced in our discussion of r_b) coincide.

Model PB entails $\mu_0=\mu_1$. Then in view of the discussion preceding Example 5, we would have $\rho=\rho_1\sqrt{pq}/\lambda=0$.

In the case of Model B the following observation is pertinent. If we consider a random sample (X_i, Z_i) in which all X observations have the same normal distribution, there will always exist chance quantities Y_i with properties:

(i) The Y_i's are independent and each has a normal distribution with mean 0 and standard deviation 1; moreover, (X_i, Y_i) is bivariate normal for each $i=1, 2, \cdots, n$

(ii) There is always a point of dichotomy ω such that the random sample (X_i, Y_i) may be dichotomized to produce the random sample (X_i, Z_i) with which we started.

This is interesting, because it means that with Model B we are never operating in a vacuum. If the X observations are normal, estimation and testing procedures will be valid for *some* ρ.

With regard to which model we should use one school of thought is the following. If one finds that the X observations are normal, and has reason to believe that the sample (X_i, Z_i) should be thought of as resulting from a dichotomy of some continuous distribution, then he may use Model B and define the ρ of the previous paragraph to be a legitimate measure of the relationship between X and Y. This is essentially the view expressed by McNemar ([7], p. 173) in relation to item evaluation for psychological tests:

"As a matter of fact, one can usually justify the use of the regular biserial r with obviously continuous variables by saying that the coefficients obtained represent what we would expect the product moment correlation to be if we had a measuring scale for the dichotomized trait which actually yielded a normal distribution. The bothersome question concerning the normality of the distribution of the trait—if and when measured in some hypothetical, unattainable, equal units—is thereby side-stepped If the point nature of the dichotomized variable can really be demonstrated, then and only then can one justify using the point-biserial correlation coefficient"

The other school of thought, and the one to which this author subscribes, may be summarized as follows One can actually observe (X_i, Z_i) It is possible that the observations may be thought of as the result of a dichotomy, but one cannot be sure. Relying on the observable characteristics of the situation one could then always consider ρ_1 to be the measure of interest, unless definite reasons can be given for assuming a continuous, normal underlying distribution. In this connection we quote Guilford ([4], p. 330), again relating to item evaluation in psychological tests.

"Since the r_{pb}, coefficient is not restricted to normal distribution in the dichotomous variable, it is much more generally applicable than is r_b. Whenever there is doubt about computing r_b, the point-biserial r will serve. For this reason it should probably be used more than it is."

In summary we offer the following recommendations.

(1) If the normality conditions are satisfied for Model PB, then use Model PB if there is any doubt concerning the validity of Model B. Employ the approximation procedures of Section 3 whenever possible. If any choice is available, try to arrange for p to be approximately $\frac{1}{2}$.

(2) If Model B is deemed to be valid, then use r_b to test hypotheses of the form $\rho=0$, and use $\hat{\rho}$, whenever it is convenient, for other purposes.

In connection with (2) it would be extremely useful to have $\sigma_{\hat{\omega}}^2$ and $\sigma_{\hat{\rho}}^2$ in tabular form. We could then not only make confidence statements about ρ using $\hat{\rho}$ but could decide how many additional observations are required to bring the accuracy of r_b up to that of $\hat{\rho}$.

REFERENCES

[1] Chernoff, H., and Lehmann, E. L., "The use of maximum likelihood estimates in chi-square tests for goodness of fit," *Annals of Mathematical Statistics*, 25 (1954), 579–86

[2] Editorial, "The normal probability function Tables of certain area-ordinate ratios and of their reciprocals," *Biometrika*, 42 (1955), 217–22.

[3] Fisher, R A., *Statistical Methods for Research Workers*. London Oliver and Boyd, 1946

[4] Guilford, J P, *Fundamental Statistics in Psychology and Education*. New York· McGraw-Hill Book Co, 1950

[5] Johnson, N. L., and Welch, B L, "Applications of the non-central t-distribution," *Biometrika*, 31 (1940), 362–89

[6] Lev, J, "The point biserial coefficient of correlation," *Annals of Mathematical Statistics*, 20 (1949), 125–6

[7] McNemar, Q, *Psychological Statistics* New York: John Wiley and Sons, 1949.

[8] Neyman, J., and Tokarska, B, "Errors of the second kind in testing 'Student's' hypothesis," *Journal of the American Statistical Association*, 31 (1936), 318–26

[9] Pearson, K, *Tables for Statisticians and Biometricians Part I*, Cambridge Cambridge University Press, 1945

[10] Soper, H E., "On the probable error for the biserial expression for the correlation coefficient," *Biometrika*, 10 (1913), 384–90.

[11] Tate, R. F., "Correlation between a discrete and a continuous variable. Point-biserial correlation," *Annals of Mathematical Statistics*, 25 (1954), 603–7

[12] Tate, R. F., "The theory of correlation between two continuous variables when one is dichotomized," *Biometrika*, 42 (1955), 205–16.

A MULTIPLE COMPARISON PROCEDURE FOR COMPARING SEVERAL TREATMENTS WITH A CONTROL

CHARLES W. DUNNETT*
American Cyanamid Company

I. INTRODUCTION

A COMMON problem in applied research is the comparison of treatments with a control or standard. Such a situation may arise, for example, when an agronomist tests the effects on crop yield of the addition of chemicals to the soil, or when a pharmacologist assays drug samples to determine their potencies. In designing an experiment to measure the effects of such treatments, it is often desirable to include in the experiment a control in the form of either a dummy treatment, to measure the magnitude of the experimental response in the absence of the treatments under investigation, or some recognized standard treatment. Sometimes past experience with the control will suffice, but often this cannot be relied upon due to altered environmental conditions. Thus the agronomist may leave a few of his experimental plots untreated for comparison with the treated plots, and the pharmacologist may measure the response to a standard drug preparation of known potency concomitantly with the test samples in order to estimate the potencies of the latter.

We will consider the case where the numerical results of an experiment performed to compare p treatments with a control can be summarized in the form of a set of numbers $\bar{X}_0, \bar{X}_1, \cdots, \bar{X}_p$ and s, where the $\bar{X}$'s are means of $p+1$ sets of observations which are assumed to be independently and normally distributed, $\bar{X}_0$ referring to the control and $\bar{X}_i$ to the i-th treatment ($i=1, \cdots, p$), and s is an independent estimate of the common standard deviation of the $p+1$ sets of observations. This paper presents a procedure for making confidence statements about the true (or expected) values of the p differences $\bar{X}_i-\bar{X}_0$, the procedure having the property that the probability of *all* p statements being simultaneously correct is equal to a specified value, P. Tables have been computed which enable the procedure to be used by the experimenter for $P=.95$ or $.99$ and $p=1(1)9$. When the numbers of observations in each set are equal, the tables enable the experimenter

* I should like to express my appreciation to Frank Wilcoxon for suggesting this problem and for his constant encouragement throughout the course of the investigation. I am grateful for this opportunity to acknowledge also my indebtedness to Robert Bechhofer and to Milton Sobel. Were it not for my good fortune to be associated with them while working on an Air Force contract at Cornell University in 1952–3, this paper could not have been written.

I am indebted to Robert E. Bechhofer also for making the tables in reference [10] available to me.

to set one-sided upper (or lower) confidence limits on the true values of the p differences $\overline{X}_i-\overline{X}_0$ such that the probability is P that all p true values will actually be less than the upper limits set, or to set two-sided confidence limits on the true values of the p differences $\overline{X}_i-\overline{X}_0$ such that P is a lower bound to the probability that all p true values will actually be between the limits set. If the numbers of observations in each set are unequal, the tables may still be used but the associated P will be only approximate. The tables may also be used to set joint confidence limits on the potencies of p drugs relative to a common standard, the associated P being approximately equal to the probability that all statements will be correct when one-sided limits are set and approximately a lower bound to this probability when two-sided limits are set.

The problem of multiple comparisons with a control is a special case of the more general multiple comparisons problem considered by Tukey [17] and Scheffé [16]. Tukey's procedure based on the Studentized range and Scheffé's procedure based on the F-distribution enable the experimenter to make any number of comparisons among a set of sample means with the assurance that the probability of all confidence statements being correct will be equal to or greater than a specified value. When the experimenter only wishes to make comparisons between one of the means and each of the others, as in the case when one of the means represents a control, use of the Tukey or Scheffé procedure would result in confidence limits which are wider than necessary. The procedure described in this paper results in narrower confidence limits for the p comparisons $\overline{X}_i-\overline{X}_0$ than either the Tukey or the Scheffé procedure

In an earlier paper, Roessler [15] considered the problem of multiple comparisons involving a control. However, he assumed that the p comparisons $\overline{X}_i-\overline{X}_0$ were independent which is incorrect since they all have $\overline{X}_0$ in common. In the present paper it is shown that, to obtain simultaneous confidence limits on the $\overline{X}_i-\overline{X}_0$, the multivariate analogue of Student's t-distribution defined by Dunnett and Sobel [4] is encountered. This same distribution was involved in a multiple decision procedure for ranking population means described by Bechhofer et al., [2], to which Tables 1a and 1b of the present paper are applicable. A multiple decision procedure for comparing several experimental categories with a control was formulated by Paulson [11]. Tables 1a and 1b of the present paper are also applicable to Paulson's procedure. The procedure described in the present paper may also be considered as a multiple decision procedure; it is compared with Paulson's in Section VII below.

For the benefit of those who may be interested primarily in applications, the procedure is illustrated by two examples in Section II. The main part of the theory is given in Section III with a description of the construction of the tables in Section IV. In Section V, the question of the optimum allocation of available experimental resources between the control and the p treatments is considered. In Section VI, the procedure is applied to the problem of estimating the potencies of p drug samples relative to a common standard.

II. EXAMPLES

(a) The following example was adapted from one given by Villars [18]. The data represent measurements on the breaking strength of fabric treated by three different chemical processes compared with a standard method of manufacture

	Breaking Strength (lbs)			
	Standard	Process 1	Process 2	Process 3
	55	55	55	50
	47	64	49	44
	48	64	52	41
Means	50	61	52	45
Variances	19	27	9	21

Here, $p=3$ and $N=3$ The average variance is $s^2=19$, which is an estimate of the common variance of the four sets with $(p+1)(N-1)=8$ degrees of freedom. It could be calculated directly as follows:

$$s^2 = \frac{55^2+47^2+48^2+55^2+\cdots+41^2-3(50^2+61^2+52^2+45^2)}{8}$$

$$= \frac{152}{8} = 19.$$

The standard deviation is $s=\sqrt{19}=4.36$ and the estimated standard error of a difference between two means is $s\sqrt{2/N}=4.36\sqrt{2/3}=3.56$ The quantity which must be added to and/or subtracted from the observed differences between the means to give their confidence limits has been called by Tukey [17] an "allowance" and is given by $A=ts\sqrt{2/N}$, where t is obtained from Table 1 if one-sided limits are desired or from Table 2 if two-sided limits are wanted. For $p=3$ and d.f. $=8$, $t=2.42$ for one-sided limits and $t=2.94$ for two-sided limits for $P=95\%$. Analogous values of t can be determined from the tables if $P=99\%$ confidence is required.

For one-sided limits, the allowance is $A=(2.42)(3.56)=9$ and the experimenter can conclude that:

(i) The breaking strength using process 1 exceeds the standard by at least $61-50-9=2$ lbs.

(ii) The breaking strength using process 2 exceeds the standard by at least $52-50-9=-7$ lbs.

(iii) The breaking strength using process 3 exceeds the standard by at least $45-50-9=-14$ lbs.

The joint statement consisting of the above three conclusions has a confidence coefficient of 95%, i.e., in the long run, 95% of such joint statements will actually be correct. Upper limits for the three differences could be obtained in an analogous manner.

For two-sided limits, the allowance is $A=(2.94)(3.56)=11$ and the experimenter can conclude that.

(i) The breaking strength using process 1 exceeds the standard by an amount between $61-50\pm 11=0$ and 22 lbs.

(ii) The breaking strength using process 2 exceeds the standard by an amount between $52-50\pm 11=-9$ and 13 lbs

(iii) The breaking strength using process 3 exceeds the standard by an amount between $45-50\pm 11=-16$ and 6 lbs.

The joint confidence coefficient for these three statements is greater than 95%. (Due to an approximation made in computing Tables 2a and 2b, the tabulated values of t are somewhat larger than necessary so that the actual P's attained are slightly greater than 95 and 99%. No such approximation was made in computing Tables 1a and 1b.)

(b) The following data are blood count measurements on three groups of animals, one of which served as a control while the other two were treated with two drugs. Due to accidental losses, the numbers of animals in the three groups are unequal

	Blood Counts (millions of cells per cubic millimeter)		
	Controls	Drug A	Drug B
	7 40	9.76	12 80
	8 50	8 80	9.68
	7.20	7.68	12.16
	8 24	9.36	9.20
	9.84		10.55
	8.32		
Sums:	49.50	35.60	54 39
N:	6	4	5
Means:	8.25	8.90	10.88

Computations:

$$s^2 = \frac{7.40^2+8.50^2+\cdots+9.20^2+10.55^2-\dfrac{49.50^2}{6}-\dfrac{35.60^2}{4}-\dfrac{54.39^2}{5}}{(\sum N_i)-(p+1)}$$

$$=\frac{16.566}{15-3}=\frac{16.566}{12}=1.3805$$

$$s=\sqrt{1.3805}=1.175$$

For d.f. $=(\sum N_i)-(p+1)=12$ and $P=95\%$, $t=2.11$ (one-sided) or $t=2.50$ (two-sided).

"Allowances" for differences from the control:

$$\textit{One-sided}: \text{drug } A: (2.11)(1.175)\sqrt{1/6+1/4} = 1.60$$

$$\text{drug } B: (2.11)(1.175)\sqrt{1/6+1/5} = 1.50$$

$$\textit{Two-sided}: \text{drug A}: (2.50)(1.175)\sqrt{1/6+1/4} = 1.90$$

$$\text{drug B}: (2.50)(1.175)\sqrt{1/6+1/5} = 1.78$$

If the experimenter is interested only in upper one-sided limits for the differences from the control, he can make the following statements:

(i) Drug A raises the blood count by at most $8.90-8.25+1.60=2.25$ millions per cmm.

(ii) Drug B raises the blood count by at most $10.88-8.25+1.50=4.13$ millions per cmm.

The joint confidence coefficient for these two statements is approximately 95% (Since the tables of t were computed for equal numbers of observations per group, their use in the case of unequal numbers results in the desired probabilities being only approximately achieved.) Corresponding lower limits could be calculated in an analogous manner.

If the experimenter desires simultaneous upper and lower limits on the differences, he should use the two-sided allowances as follows:

(i) Drug A raises the blood count by an amount between $8.90-8.25\pm1.90=-0.25$ and 2.55 millions per cmm.

(ii) Drug B raises the blood count by an amount between $10.88-8.25\pm1.78=0.85$ and 4.41 millions per cmm.

An approximate lower bound to the joint confidence coefficient of these statements is 95%.

III. THEORETICAL BASIS

Suppose there are available N_0 observations on the control, N_1 observations on the first treatment, $\cdots$, N_p observations on the p-th treatment. Denote these observations by X_{ij} ($i=0, 1, \cdots, p$; $j=1, 2, \cdots, N_i$), and the i-th treatment mean, $\sum_{j=1}^{N_i} X_{ij}/N_i$, by $\overline{X}_i$. We make the assumptions usually made in the analysis of variance, namely, that the X_{ij} are independent and normally distributed with common variance σ^2 and means m_i. We assume also that there is available an estimate s^2 of σ^2, independent of the $\overline{X}_i$, which is based on n degrees of freedom. For example, we may take

$$s^2 = \sum_{i=0}^{p} \sum_{j=1}^{N_i} (X_{ij} - \overline{X}_i)^2/n \tag{1}$$

where $n = (\sum_{i=0}^{p} N_i) - (p+1)$ Our problem is to obtain separate confidence limits for each of the differences $m_i - m_0$ ($i=1, 2, \cdots, p$), such that the joint confidence coefficient, i e, the probability that all p confidence intervals will contain the corresponding $m_i - m_0$, is equal to a preassigned value, P $(0<P<1)$.

Consider first the case $p=1$, where there is only one treatment to be compared with the control The method of obtaining confidence limits for $m_1 - m_0$, as described in almost any statistics textbook, is based on Student's t-distribution If we write

$$z = \frac{\overline{X}_1 - \overline{X}_0 - (m_1 - m_0)}{\sqrt{\dfrac{1}{N_1} + \dfrac{1}{N_0}}}$$

which is normally distributed with mean 0 and variance σ^2, then $t = z/s$ follows the Student t-distribution with n degrees of freedom. A lower limit on $m_1 - m_0$ with the desired confidence coefficient will be given by

$$\overline{X}_1 - \overline{X}_0 - d's\sqrt{\frac{1}{N_1} + \frac{1}{N_0}}$$

if d' is chosen so that

$$\text{Prob}\,(t < d') = P\ . \tag{2}$$

Similarly, if an upper confidence limit is required, it will be given by

$$\overline{X}_1 - \overline{X}_0 + d's\sqrt{\frac{1}{N_1} + \frac{1}{N_0}}\ .$$

On the other hand, if the experimenter wishes to have bounds on $m_1 - m_0$ in both directions, he may take

$$\overline{X}_1 - \overline{X}_0 \pm d''s\sqrt{\frac{1}{N_1} + \frac{1}{N_0}}$$

where d'' is chosen to satisfy

$$\text{Prob}\ (|t| < d'') = P \tag{3}$$

The constant d' or d'' corresponding to the desired value of P can be obtained from tables of the percentage points of Student's t-distribution, which are widely available

Now consider the general case where there are p treatments and a control Write

$$z_i = \frac{\overline{X}_i - \overline{X}_0 - (m_i - m_0)}{\sqrt{\frac{1}{N_i} + \frac{1}{N_0}}}$$

and $t_i = z_i/s$, $i = 1, 2, \cdots, p$. Then lower confidence limits with joint confidence coefficient P for the p treatment effects $m_i - m_0$ will be given by

$$\overline{X}_i - \overline{X}_0 - d_i's\sqrt{\frac{1}{N_i} + \frac{1}{N_0}}, \quad (i = 1, 2, \cdots, p),$$

if the p constants d_i' are chosen so that

$$\text{Prob}\ (t_1 < d_1', t_2 < d_2', \cdots, t_p < d_p') = P. \tag{4}$$

Similarly, upper confidence limits will be given by

$$\overline{X}_i - \overline{X}_0 + d_i's\sqrt{\frac{1}{N_i} + \frac{1}{N_0}}$$

On the other hand, two-sided confidence limits having the desired joint confidence coefficient will be given by

$$\overline{X}_i - \overline{X}_0 \pm d_i''s\sqrt{\frac{1}{N_i} + \frac{1}{N_0}} \quad (i = 1, 2, \cdots, p),$$

if the p constants d_i'' are chosen to satisfy

$$\text{Prob}\ (|t_1| < d_1'', |t_2| < d_2'', \cdots, |t_p| < d_p'') = P. \tag{5}$$

To find any set of constants d_i' or d_i'' satisfying these equations, the joint distribution of the t_i is required. It may be noted that the joint distribution of the z_i is a multivariate normal distribution with means 0 and variances σ^2 and where the correlation between z_i and z_j is given by

$$\rho_{ij} = 1/\sqrt{\left(\frac{N_0}{N_i}+1\right)\left(\frac{N_0}{N_j}+1\right)}.$$

The joint distribution of the t_i is thus the multivariate analogue of Student's t-distribution defined by Dunnett and Sobel [4].

To find solutions to (4) and (5), we require a tabulation of the multivariate Student t-distribution. We shall show now that the problem of tabulating the multivariate Student t-distribution can be reduced to the problem of tabulating the corresponding multivariate normal distribution. Consider first equation (4) above. It can be written

$$\begin{aligned} P &= \text{Prob}\,(z_1 < d_1's, z_2 < d_2's, \cdots, z_p < d_p's) \\ &= \int_{-\infty}^{+\infty} F(d_1's, d_2's, \cdots, d_p's)p(s)ds \end{aligned} \tag{6}$$

where $F(z_1, z_2, \cdots, z_p)$ is the multivariate normal c d f of the z_i and $p(s)$ is the probability density function of s. Thus, if F were tabulated, it would be fairly easy using a desk calculator to evaluate (6) by numerical integration for any set of fixed d_i', and hence to find solutions to (4) as functions of P, p and n.

Similarly, (5) can be written

$$\begin{aligned} P &= \text{Prob}\,(|z_1| < d_1''s, |z_2| < d_2''s, \cdots, |z_p| < d_p''s) \\ &= \int_{-\infty}^{+\infty} G(d_1''s, d_2''s, \cdots, d_p''s)p(s)ds \end{aligned} \tag{7}$$

where $G(z_1, z_2, \cdots, z_p)$ is the c d f of the $|z_i|$ Again, if G were tabulated, we could also evaluate (7) and determine the solutions to (5) as functions of P, p and n.

The functions F and G can be obtained for $p=2$ from K Pearson's tables of the bivariate normal distribution [13], although the tabulation interval is not fine enough for numerical integration purposes. However, for $p=2$, (4) and (5) can be evaluated directly from the results of Dunnett and Sobel [4]. The function F has been tabulated for equal values of its arguments and for $p \leq 9$ by the National Bureau of Standards [10] for the special case $\rho_{ij}=1/2$. As will be explained in the next

section, Table 1 of this paper was based on this tabulation. It would be extremely useful for the purpose of the problem considered in this paper to have a corresponding tabulation of the function G.

Until exact tables are available, it will be necessary in general to rely upon approximations to the solutions of (4) and (5). From the results given by Dunnett and Sobel [5], we can write, when

$$\rho_{ij} = 1 \Big/ \sqrt{\left(\frac{N_0}{N_i} + 1\right)\left(\frac{N_0}{N_j} + 1\right)},$$

$$\text{Prob}\,(t_1 < d_1', t_2 < d_2', \cdots, t_p < d_p') \geqq \prod_{i=1}^{p} \text{Prob}\,(t < d_i') \tag{8}$$

and

$$\begin{aligned} &\text{Prob}\,(|t_1| < d_1'', |t_2| < d_2'', \cdots, |t_p| < d_p'') \\ &\qquad \geqq \prod_{i=1}^{p} \text{Prob}\,(|t| < d_i''). \end{aligned} \tag{9}$$

Thus, upper bounds to the constants d_i' and d_i'' can be determined by equating the right-hand sides of (8) and (9) to the desired value of P. These calculations involve only the probability integral of the univariate Student t-distribution, which has been tabulated most recently by Hartley and Pearson [8, 12].

Since (4) and (5) are each increasing functions of the correlations ρ_{ij}, alternative lower bounds which are closer than the lower bounds given in (8) and (9) may be obtained by taking $\rho_{ij}=0$. Pillai and Ramachandran [14] have tabulated solutions $d_1'=d_2'= \cdots d_p'$ to (4) and solutions $d_1''=d_2''= \cdots =d_p''$ to (5) for $P=.95$ and $\rho_{ij}=0$. It would be useful to have a tabulation of such equicoordinate percentage points of the multivariate Student t-distribution in the important special case where $\rho_{ij}=0$ for other values of P.

Lower bounds based on the bivariate Student t-distribution can also be obtained, as shown in [5]. Taking $d_1'=d_2'= \cdots =d_p'=d'$, $d_1''=d_2''= \cdots =d_p''=d''$, and $\rho_{ij}=\rho\geqq 0$, these can be written

$$\begin{aligned} &\text{Prob}\,(t_1 < d', t_2 < d', \cdots, t_p < d' \mid \rho_{ij} = \rho) \\ &\qquad \geqq [\text{Prob}\,(t_1 < d', t_2 < d' \mid \rho_{12} = \rho)]^{p/2} \end{aligned} \tag{10}$$

and

$$\begin{aligned} &\text{Prob}\,(|t_1| < d'', |t_2| < d'', \cdots, |t_p| < d'' \mid \rho_{ij} = \rho) \\ &\qquad \geqq [\text{Prob}\,(|t_1| < d'', |t_2| < d'' \mid \rho_{12} = \rho)]^{p/2} \end{aligned} \tag{11}$$

The probability within the square brackets can be determined from the probability integral of the bivariate Student-t distribution [4] with correlation ρ. The bounds obtained from these inequalities are sharper than those given by (8) and (9), and in most cases will also be sharper than those obtained by using the Pillai and Ramachandran tables [14].

There are, of course, infinitely many solutions to (4) and (5). For the applications considered in this paper, we will take the constants to be equal, viz., $d_1'=d_2'= \cdots =d_p'$ ($=d'$, say) and $d_1''=d_2''= \cdots =d_p''$ ($=d''$, say). Besides greatly simplifying the computational problem, there are some theoretical grounds for doing so in the case where $\rho_{ij}=\rho \geq 0$, which occurs frequently in practice. For example, it can be shown that $\sum d_i'^2$ and $\sum d_i''^2$ are each minimized by this choice.

IV. CONSTRUCTION OF THE TABLES

Tables 1a and 1b give solutions $d'=d_1'= \cdots =d_p'$ to (4) for $P=.95$ and .99, respectively, for $p \leq 9$ and $\rho_{ij}=1/2$. They are applicable to the situation where there are equal numbers of observations on the control and the p treatments, viz, $N_0=N_1= \cdots =N_p$. These tables were constructed by numerical evaluation of the integral in (6) using tables of the function F computed by the National Bureau of Standards [10] The method used was as follows. For $n=5$, 10, 20, 25 degrees of freedom, the integral in (6) was calculated for three successive values of d' differing by 0.1 such that the desired value of P was bracketed. The required value of d' was then determined to 3 decimal places by 3-point inverse interpolation. For $n=\infty$ degrees of freedom, the values given in tables computed by Bechhofer [1] were used For the intermediate degrees of freedom, the values were obtained by interpolation with $1/n$ as argument An accuracy of 1 in the second decimal place should be achieved by this method.

Tables 2a and 2b give solutions $d''=d_1''= \cdots =d_p''$ to the right-hand side of (11) for $P=.95$ and .99, respectively, for $p \leq 9$ and $\rho_{ij}=1/2$. The method of constructing these tables was as follows: For $n=5$, 10, 20, 40, ∞ degrees of freedom, the probability in the square brackets of (11) was calculated for three successive values of d'' differing by 0.1 such that the desired value of P was bracketed, using the expressions developed in [4] for the probability integral of the bivariate Student t-distribution. The required value of d'' was then determined by 3-point inverse interpolation to 3 decimal places. For the intermediate degrees of freedom, the required values were obtained by interpolation using $1/n$ as argument. An accuracy of 1 in the second decimal place should be achieved by this method.

V. OPTIMUM ALLOCATION OF OBSERVATIONS BETWEEN CONTROL AND TREATMENTS

The tables given in this paper were prepared to handle the case where equal numbers of observations are available on each of the p treatments and on the control. In many practical situations, the experimenter will, in fact, wish to allocate the available number of observations equally to each group. Where it is feasible, however, it may be advantageous to do otherwise. In this section, we will consider the consequences of allocating N_0 observations to the control and N_1 observations to each of the other treatment groups, i.e., we will take $N_1 = N_2 = \cdots = N_p$ and $N_0 \neq N_1$

Lower confidence limits to the $m_i - m_0$ will be given by

$$\overline{X}_i - \overline{X}_0 - d's\sqrt{\frac{1}{N_1} + \frac{1}{N_0}} \qquad (i = 1, 2, \cdots, p),$$

where d' is chosen to satisfy

$$\text{Prob } (t_i < d', i = 1, 2, \cdots, p) = P, \tag{12}$$

the t_i having the multivariate Student t-distribution with

$$\rho_{ij} = 1 \bigg/ \left(\frac{N_0}{N_1} + 1\right).$$

We will consider the allocation N_0/N_1 optimum if it maximizes P for fixed value of

$$d'\sqrt{\frac{1}{N_1} + \frac{1}{N_0}}$$

and fixed total number of observations, $N_0 + pN_1$. It may be noted that fixing the total number of observations also fixes n, the number of degrees of freedom associated with the multivariate t-distribution, if s is defined by (1). Let

$$h = d'\sqrt{\frac{1}{N_1} + \frac{1}{N_0}} \cdot$$

Then (12) can be written

$$\text{Prob}\left[t_i < \frac{h}{\sqrt{\frac{1}{N_1} + \frac{1}{N_0}}}, i = 1, 2, \cdots, p\right] = P \tag{13}$$

From (13), it is evident that P can be increased for fixed h by decreasing

$$\sqrt{\frac{1}{N_1}+\frac{1}{N_0}}\,.$$

It is easy to show, as has been pointed out by several authors, see, for example, Finney [7], that

$$\sqrt{\frac{1}{N_1}+\frac{1}{N_0}}$$

attains a minimum for fixed N_0+pN_1 when $N_0/N_1=\sqrt{p}$. However, this choice makes $\rho_{ij}<1/2$, which operates to decrease P

In order to investigate numerically the effect of different allocations on P, the curves shown in Figs 1 and 2 computed. Fig. 1 shows P as a function of N_0/N_1 for $p=2$ and $n=1, 2, 5, 10, \infty$, with h chosen in each case to make $P=.95$ when $N_0/N_1=1$ Fig 2 shows P as a function of N_0/N_1 for $p=2, 4, 9$ and $n=\infty$, with h again chosen to make $P=.95$ when $N_0/N_1=1$ The curves for finite n were computed from formulas given in [4] The curves for $n=\infty$ were computed by numerical integration using tables of the normal distribution [9], with certain points for $p=2$ checked against Pearson's bivariate normal tables [13].

Figs 1 and 2 indicate that the optimum allocation occurs where N_0/N_1 is only slightly less than $\sqrt{p}$ except when the number of degrees of freedom is small Some further computations carried out by the author for the case of $n=\infty$ degrees of freedom indicated that the point of optimum allocation becomes even closer to $\sqrt{p}$ when h is chosen to make $P=.99$ for $N_0/N_1=1$. However, for h chosen to make $P=.75$ for $N_0/N_1=1$, it was found that $P<.75$ for $N_0/N_1=\sqrt{p}$ and the optimum value of N_0/N_1 was considerably less than $\sqrt{p}$. For practical purposes, we can thus conclude that, if the experimenter is working with a joint confidence coefficient in the neighborhood of $P=.95$ or greater, then the experiment should be designed so that $N_0/N_1=\sqrt{p}$ approximately, where N_0 is the number of observations on the control and N_1 the number on each of the p treatments

VI. APPLICATION TO BIOLOGICAL ASSAY

An important example involving the multiple comparison of several treatments with a control arises in the biological assay of several drug samples relative to a common standard. In the "parallel line" type of biological assay, regression lines $Y=a_0+bX$ and $Y=a_1+bX$ are fitted

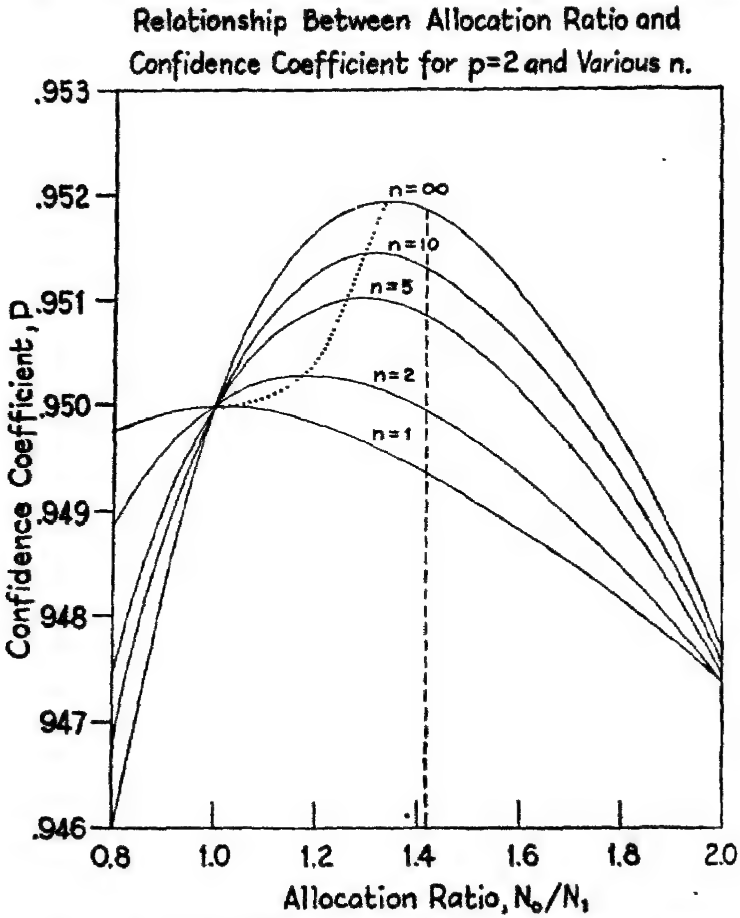


FIG. 1.—In plotting these curves, $d'\sqrt{(1/N_1)+(1/N_0)}$ was fixed for each n so that $P=.95$ for $N_0=N_1$, where $d's\sqrt{(1/N_1)+(1/N_0)}$ is the difference between $\bar{X}_1-\bar{X}_0$ and its lower confidence limit. The dotted curve indicates where the optimum occurs for each n; the vertical dashed line is drawn at $N_0/N_1=\sqrt{p}=\sqrt{2}$

to data representing observed responses Y at several log-dose levels X of a standard drug preparation S and a test sample U. The estimated log-potency of U relative to S is represented by the horizontal distance between the two regression lines,

$$M = \frac{a_1 - a_0}{b} = \bar{X}_0 - \bar{X}_1 + \frac{\bar{Y}_1 - \bar{Y}_0}{b} \tag{14}$$

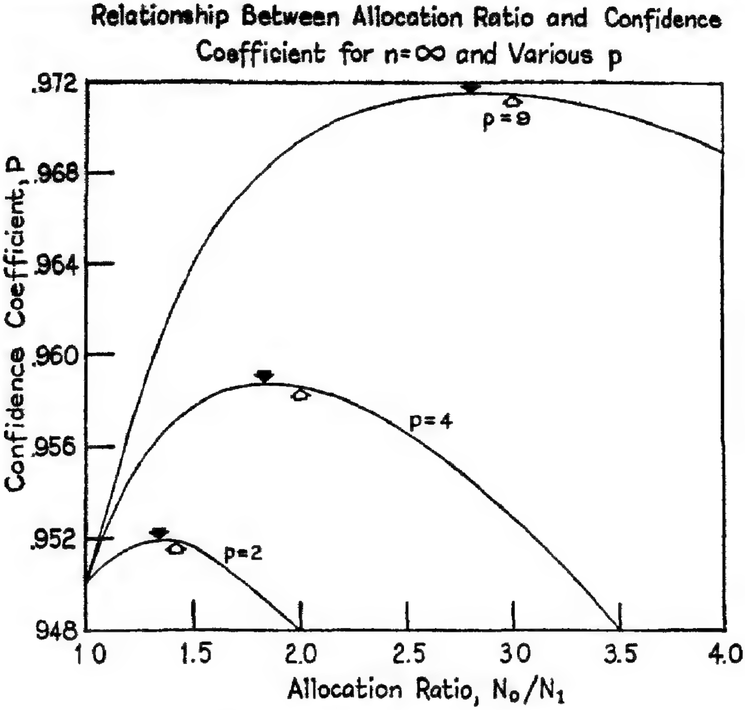


FIG. 2.—In plotting these curves, $d'\sqrt{(1/N_1)+(1/N_0)}$ was fixed for each p so that $P=.95$ for $N_0=N_1$, where $d'\sigma\sqrt{(1/N_1)+(1/N_0)}$ is the difference between $\bar{X}_i-\bar{X}_0$ and its lower confidence limit. The black arrows indicate where the optima occur; the white arrows indicate the points where $N_0/N_1=\sqrt{p}$.

where $\bar{X}_0$ and $\bar{X}_1$ are the mean log-dose levels, and $\bar{Y}_0$ and $\bar{Y}_1$ are the mean observed responses, of S and U respectively. On the other hand, the true log-potency may be represented by

$$\bar{M} = \bar{X}_0 - \bar{X}_1 + \frac{m_1 - m_0}{\beta} \tag{15}$$

where m_0, m_1 and β are the expected values of $\bar{Y}_0$, $\bar{Y}_1$ and b, respectively Assuming that the responses Y are independent and normally distributed with common variance σ^2 and means $m_0+\beta(X-\bar{X}_0)$ and $m_1+\beta(X-\bar{X}_1)$ for S and U respectively, we can obtain confidence limits for $\bar{M}$ by Fieller's [6] method by considering

$$z = \frac{\bar{Y}_1 - \bar{Y}_0 - (\bar{M} - \bar{X}_0 + \bar{X}_1)b}{\left[\frac{1}{N_1} + \frac{1}{N_0} + (\bar{M} - \bar{X}_0 + \bar{X}_1)^2 c^2\right]^{1/2}} \tag{16}$$

Here, N_0 and N_1 are the numbers of responses observed on S and U respectively, and $c^2\sigma^2$ is the variance of the common slope b Then z is normally distributed with mean 0 and variance σ^2 If s^2 is an estimate of σ^2 independent of z, based on n degrees of freedom, then z/s has Student's t-distribution with n degrees of freedom. Hence, confidence limits for $\bar{M}$ with confidence coefficient P can be obtained by equating z^2/s^2 to t^2, where t is the P-percentage point of Student's t, and solving the resulting quadratic equation for $\bar{M}$. The solution may be found, for example, in Bliss [3].

Now suppose there are p unknown drug samples: $U_1, U_2 \cdots, U_p$, to be compared with a common standard, S. Extending the above notation in an obvious way, we define the p variables,

$$z_i = \frac{\bar{Y}_i - \bar{Y}_0 - (\bar{M}_i - \bar{X}_0 + \bar{X}_i)b}{\left[\frac{1}{N_i} + \frac{1}{N_0} + (\bar{M}_i - \bar{X}_0 + \bar{X}_i)^2 c^2\right]^{1/2}}, \quad i = 1, 2, \cdots, p \tag{17}$$

The z_i have a joint p-variate normal distribution with means 0, common variance σ^2 and correlation between z_i and z_j given by

$$\rho_{ij} = (1 + \epsilon_i\epsilon_j) \Big/ \sqrt{\left(\frac{N_0}{N_i} + 1 + \epsilon_i^2\right)\left(\frac{N_0}{N_j} + 1 + \epsilon_j^2\right)}$$

where $\epsilon_i = \bar{M}_i - \bar{X}_0 + \bar{X}_i)c\sqrt{N_0}$. For the p pairs of confidence limits on the $\bar{M}_i$ to have a joint confidence coefficient equal to P, we are led to consider the joint distribution of the z_i/s, which is the multivariate Student t-distribution as in Section III, with this important difference. the correlations ρ_{ij} are no longer known exactly since they involve the unknown parameters $\bar{M}_i$ and $\bar{M}_j$ Fortunately, ϵ_i may be expected to be fairly small in most cases since it is common practice in designing a biological assay experiment to try to arrange the dose levels so that $\bar{X}_0 - \bar{X}_i$ is close to $\bar{M}_i$. Thus, we can obtain confidence limits for the $\bar{M}_i$ which have approximately the desired confidence coefficient by assuming the ϵ_i to be negligible, whence

$$\rho_{ij} = 1 \Big/ \sqrt{\left(\frac{N_0}{N_i} + 1\right)\left(\frac{N_0}{N_j} + 1\right)}. \tag{18}$$

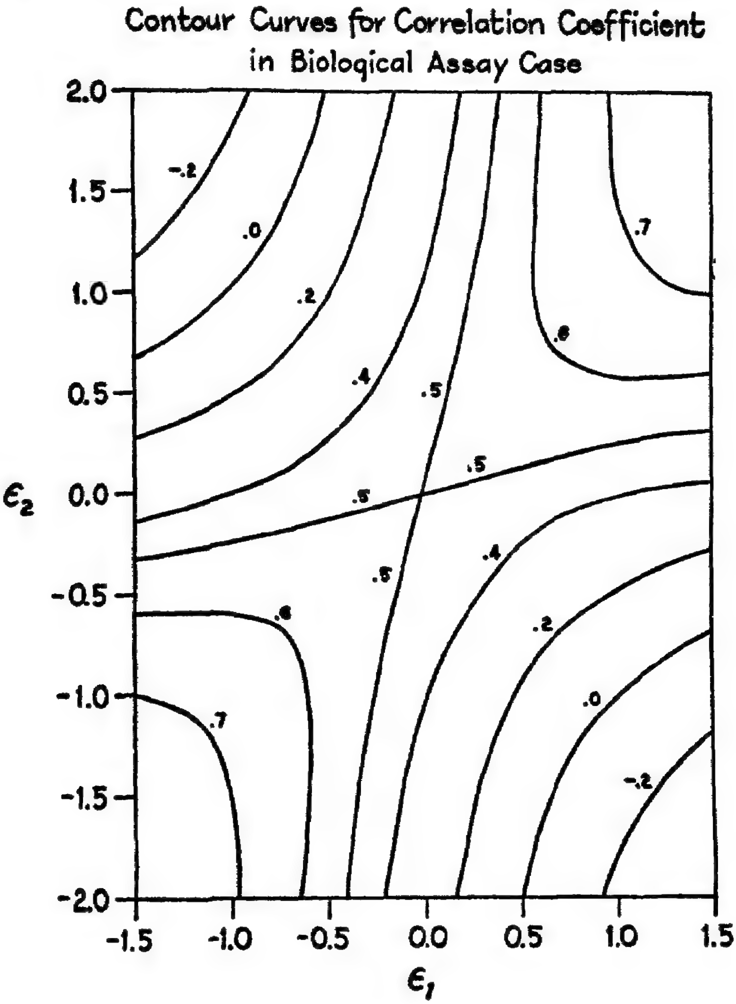


FIG. 3

This takes the value $\frac{1}{2}$ in the usual situation where $N_0 = N_i$ $(i=1, 2, \cdots, p)$, in which case the tables in this paper are applicable.

The contour curves in Fig 3 show the effect of ϵ_i and ϵ_j on ρ_{ij} when $N_0 = N_i = N_j$ In most practical situations, the log-dose levels chosen by the experimenter for each drug will cover a wide enough range so that

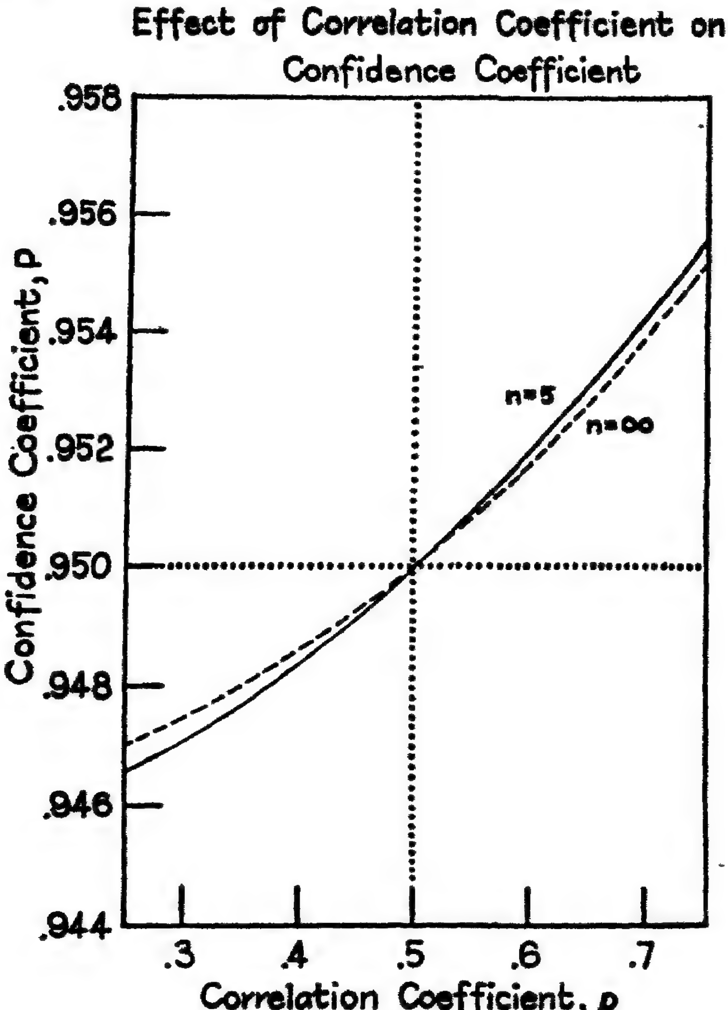


FIG. 4

the ϵ_i are numerically small, in which case they do not exert much influence on the value of ρ_{ij}. In Fig. 4, curves are drawn for degrees of freedom equal to 5 and ∞ showing the actual P attained as a function of ρ in the case of estimating the log-potencies of two drugs ($p=2$) when $N_0=N_1=N_2$ and the tables are used applicable to $\rho=\frac{1}{2}$ and

$P=95\%$. It may be seen that for a wide range of values of ρ around $\rho=\frac{1}{2}$, P does not differ much from the value 95%.

VII. MULTIPLE DECISION PROCEDURES FOR COMPARING TREATMENTS WITH A CONTROL

A multiple decision procedure for selecting the "best" of $p+1$ categories when comparing p experimental categories with a control has been developed by Paulson [11]. Paulson gives a method for making one of the following $p+1$ decisions:

$$\begin{cases} D_0\text{: select the control as best} \\ D_i\text{: select the } i\text{-th category as best} \qquad (i=1, 2, \cdots, p) \end{cases}$$

If $\overline{X}_0, \overline{X}_1, \cdots, \overline{X}_p$ represent the $p+1$ means, each based on N observations, $\overline{X}_0$ being the control, then Paulson's method consists in picking out $\overline{X}^*=\max(\overline{X}_1, \cdots, \overline{X}_p)$ and if $\overline{X}^*-\overline{X}_0\geqq\lambda_n s\sqrt{2/N}$ the decision D^* corresponding to $\overline{X}^*$ is made, whereas if $\overline{X}^*-\overline{X}_0<\lambda_n s\sqrt{2/N}$ the decision D_0 is made The constant λ_n is chosen so that the probability of making decision D_0, when all p treatments are equivalent to the control, is equal to a pre-assigned value P. Clearly, $d_t'=\lambda_n$ must be a solution of (4). Thus, Tables 1a and 1b of this paper give the required values of λ_n for $P=.95$ and $.99$.

In some situations it may be appropriate to consider the following set of possible decisions from which a choice is to be made:

$$\begin{cases} D_0\text{: select the control as best} \\ D_i\text{: select the } i\text{-th treatment } (i=1, 2, \cdots, p) \text{ as the only one better than the control} \\ D_{ij}\text{: select the } i\text{-th and } j\text{-th treatments } (i, j=1, 2, \cdots, p,\ i\neq j)\text{, without ordering them, as the only two better than the control} \\ \cdots \\ D_{1,2,\cdots,p}\text{: select all } p \text{ treatments, without ordering them, as better than the control} \end{cases}$$

The following procedure is proposed for choosing one of the above 2^p decisions on the basis of the $p+1$ observed means $\overline{X}_0, \overline{X}_1, \cdots, \overline{X}_p$, each based on N observations, and the independent estimate s of the common standard deviation:

$$\begin{cases} \text{Accept } D_0 \text{ if } \overline{X}_i-\overline{X}_0<ds\sqrt{2/N} \text{ for all } i \\ \text{Accept } D_i \text{ if } \overline{X}_i-\overline{X}_0\geqq ds\sqrt{2/N} \text{ and } \overline{X}_j-\overline{X}_0<ds\sqrt{2/N} \text{ for all } j\neq i \\ \text{Accept } D_{ij} \text{ if } \overline{X}_i-\overline{X}_0\geqq ds\sqrt{2/N},\ \overline{X}_j-\overline{X}_0\geqq ds\sqrt{2/N} \text{ and } \overline{X}_k-\overline{X}_0 <ds\sqrt{2/N} \text{ for all } k\neq i \text{ or } j \\ \cdots \\ \text{Accept } D_{1,2,\cdots,p} \text{ if } \overline{X}_i-\overline{X}_0\geqq ds\sqrt{2/N} \text{ for all } i \end{cases}$$

By choosing d to satisfy (4), we will be assured that the probability of accepting D_0 when all the treatments are equivalent to the control is equal to a pre-assigned value P. In fact, d will then be identical with Paulson's λ_n. Table 1 gives the required values of d for $P=.95$ and $.99$.

It is also possible to determine the size N of sample required to achieve a specified probability of accepting some other decision when all the treatments are not equivalent to the control. For example, suppose all the treatments are equivalent to the control except one, say the first one, which is better than the control. The correct decision to make would then be D_1. Suppose $m_0=m_2=m_3=\cdots=m_p=m_1-\delta$. Then the probability of making decision D_1 can be written

$$\begin{aligned} P_1 &= \Pr\,(D_1 \mid m_0 = m_2 = \cdots = m_p = m_1 - \delta) \\ &= \Pr\,(\overline{X}_1 - \overline{X}_0 \geqq ds\sqrt{2/N},\quad \overline{X}_2 - \overline{X}_0 \\ &< ds\sqrt{2/N}, \cdots, \overline{X}_p - \overline{X}_0 < ds\sqrt{2/N}) \\ &= \Pr\,(t_1 \geqq d, t_2 < d, \cdots, t_p < d) \end{aligned} \tag{19}$$

where

$$t_i = \frac{\overline{X}_i - \overline{X}_0}{s\sqrt{2/N}} \qquad (i = 1, 2, \cdots, p)$$

We note that $t_2, \cdots, t_p$ are Student t-variates, but t_1 is a non-central t-variate.

To obtain bounds for P_1, write

$$\begin{aligned} \overline{P}_1 &= \Pr\,(t_1 < d, t_2 < d, \cdots, t_p < d) \\ &\geqq \Pr\,(t_1 < d)\ \Pr\,(t_2 < d, \cdots, t_p < d) \end{aligned} \tag{20}$$

This inequality follows from Dunnett and Sobel [5]. Since

$$P_1 + \overline{P}_1 = \Pr(t_2 < d, \cdots, t_p < d), \tag{21}$$

it follows that an upper bound for P_1 is given by

$$P_1 \leqq \Pr\,(t_1 \geqq d)\cdot \Pr\,(t_2 < d, \cdots, t_p < d). \tag{22}$$

On the other hand, an obvious upper bound on $\overline{P}_1$ is

$$\overline{P}_1 \leqq \Pr\,(t_1 < d) \tag{23}$$

From (21) and (23), we get the following lower bound for P_1,

$$P_1 \geqq \Pr\,(t_2 < d, \cdots, t_p < d) - \Pr\,(t_1 < d). \tag{24}$$

The difference between the bounds given by (22) and (24) is

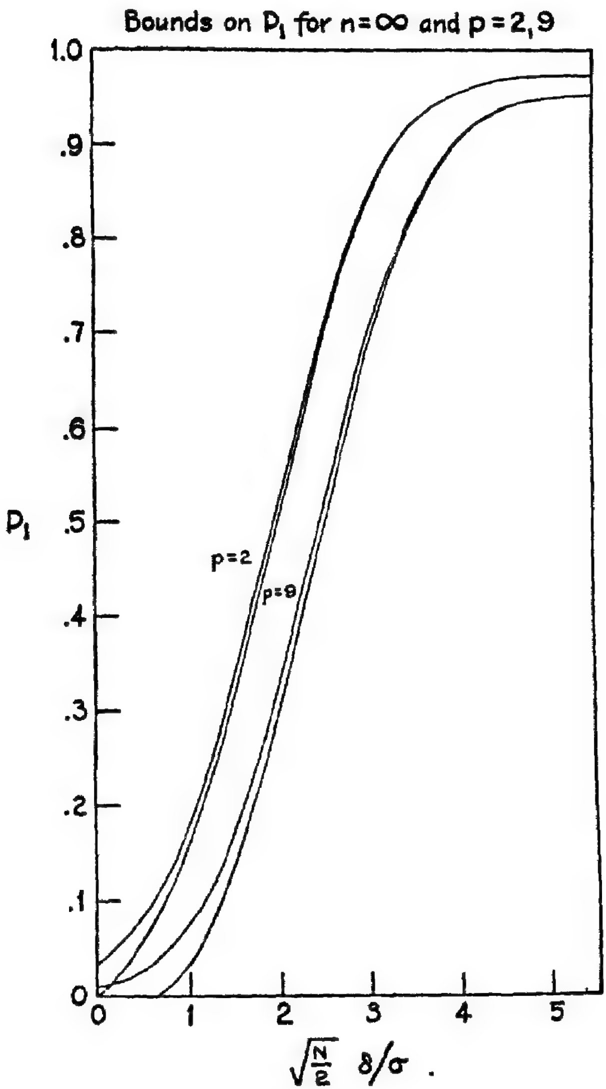

Bounds on P_1 for $n=\infty$ and $p=2,9$
1.0
.9
.8
.7
.6
P_1
.5
.4
.3
.2
.1
0
p=2
p=9
0
1
2
3
4
5
$\sqrt{\frac{N}{2}}\ \delta/\sigma$.

FIG. 5

$\Pr(t_1<d)\cdot[1-\Pr(t_2<d, \cdots, t_p<d)]$ which will be small in the region of interest where P_1 is fairly large. In Fig. 5, the two bounds for P_1 are plotted as functions of $\sqrt{(N/2)}\delta/\sigma$ for d.f. $=\infty$, $p=2$ and 9, and where d is chosen in each case so that $P=.95$. To compute the probability involving $t_2, \cdots, t_p$, in (22) and (24), tables of the multivariate normal distribution computed by the Bureau of Standards [10] were used. If it were desired to compute the bounds on P_1 for a finite number of degrees of freedom, the probability involving $t_2, \cdots, t_p$ could be computed by numerical integration on the Bureau of Standards tables. However, since in most practical situations the number of degrees of freedom will be large, it may be sufficient to assume d.f. $=\infty$ and use the Bureau of Standards tables directly.

Using Fig. 5, it is possible to determine the required value of N corresponding to any given values of δ/σ and P_1. The following table shows the value of $\sqrt{N}\delta/\sigma$ corresponding to $p=1(1)9$ to achieve $P_1=.80$ for $P=.95$ and d.f $=\infty$.

p	$\sqrt{N}\delta/\sigma$
1	3 52
2	4.05
3	4.30
4	4.46
5	4.58
6	4 67
7	4.74
8	4.81
9	4.86

For example, suppose that the experimenter has $p=5$ treatments to compare with a control. Table 1a will provide the value of d to give a probability of $P=.95$ that none of the treatments will be declared superior to the control when, in fact, all are equivalent to the control. If, in addition, the experimenter wants to achieve a probability of $P_1=.80$ of correctly selecting a superior treatment, if there is one which is, say, one standard deviation better than the control, the above table shows that $\sqrt{N}\delta/\sigma=4.58$ whence $N=21$ on substituting $\delta/\sigma=1$. Thus 21 observations should be taken on each of the treatments and on the control. This will provide $(p+1)(N-1)=120$ degrees of freedom for estimating the variance, so that $d=2.26$ from table 1a. While the table used above to determine N is based on d.f. $=\infty$, the result should not be much different for d.f. $=120$. The experimenter would then take 21 ob-

servations in each group, observe $\overline{X}_0$ and $\overline{X}_i$ ($i=1, 2, \cdots, 5$), calculate s^2 from equation (1), and declare any treatment superior to the control which gives a mean $\overline{X}_i$ greater than $\overline{X}_0+2.26s\sqrt{2/21}$.

TABLE 1a*

TABLE OF t FOR ONE-SIDED COMPARISONS BETWEEN p TREATMENT MEANS AND A CONTROL FOR A JOINT CONFIDENCE COEFFICIENT OF $P=95\%$

	p, Number Of Treatment Means (Excluding The Control)								
d f.	1	2	3	4	5	6	7	8	9
5	2.02	2 44	2.68	2 85	2.98	3.08	3 16	3 24	3 30
6	1.94	2 34	2 56	2 71	2 83	2.92	3 00	3.07	3.12
7	1.89	2.27	2.48	2 62	2 73	2 82	2 89	2 95	3 01
8	1.86	2 22	2 42	2.55	2 66	2 74	2.81	2.87	2 92
9	1 83	2.18	2.37	2 50	2 60	2.68	2 75	2 81	2 86
10	1 81	2.15	2 34	2 47	2.56	2 64	2 70	2.76	2.81
11	1.80	2 13	2 31	2 44	2 53	2.60	2.67	2 72	2 77
12	1.78	2 11	2 29	2.41	2 50	2 58	2 64	2 69	2.74
13	1.77	2 09	2.27	2 39	2.48	2 55	2 61	2 66	2 71
14	1 76	2.08	2 25	2.37	2 46	2.53	2.59	2.64	2 69
15	1 75	2 07	2 24	2 36	2 44	2.51	2.57	2 62	2 67
16	1.75	2 06	2 23	2.34	2 43	2.50	2 56	2.61	2.65
17	1 74	2.05	2 22	2.33	2.42	2.49	2.54	2 59	2.64
18	1 73	2 04	2 21	2.32	2.41	2.48	2 53	2 58	2 62
19	1.73	2 03	2 20	2 31	2.40	2 47	2 52	2 57	2.61
20	1.72	2 03	2.19	2 30	2 39	2 46	2.51	2 56	2 60
24	1.71	2 01	2.17	2 28	2 36	2 43	2 48	2 53	2.57
30	1.70	1 99	2 15	2.25	2.33	2 40	2.45	2 50	2 54
40	1.68	1 97	2 13	2 23	2.31	2.37	2.42	2 47	2 51
60	1.67	1.95	2 10	2 21	2 28	2.35	2.39	2.44	2.48
120	1.66	1 93	2 08	2 18	2.26	2.32	2.37	2 41	2 45
inf.	1.64	1.92	2.06	2.16	2 23	2.29	2 34	2 38	2 42

* Table 1a gives a solution $d_i'=t$ to equation (4) in the text for $P=.95$ for the case $\rho_{ij}=1/2$.

TABLE 1b*

TABLE OF t FOR ONE-SIDED COMPARISONS BETWEEN p TREATMENT MEANS AND A CONTROL FOR A JOINT CONFIDENCE COEFFICIENT OF $P=99\%$

	p, Number Of Treatment Means (Excluding The Control)								
d.f.	1	2	3	4	5	6	7	8	9
5	3.37	3.90	4.21	4.43	4.60	4.73	4.85	4.94	5.03
6	3.14	3.61	3.88	4.07	4.21	4.33	4.43	4.51	4.59
7	3.00	3.42	3.66	3.83	3.96	4.07	4.15	4.23	4.30
8	2.90	3.29	3.51	3.67	3.79	3.88	3.96	4.03	4.09
9	2.82	3.19	3.40	3.55	3.66	3.75	3.82	3.89	3.94
10	2.76	3.11	3.31	3.45	3.56	3.64	3.71	3.78	3.83
11	2.72	3.06	3.25	3.38	3.48	3.56	3.63	3.69	3.74
12	2.68	3.01	3.19	3.32	3.42	3.50	3.56	3.62	3.67
13	2.65	2.97	3.15	3.27	3.37	3.44	3.51	3.56	3.61
14	2.62	2.94	3.11	3.23	3.32	3.40	3.46	3.51	3.56
15	2.60	2.91	3.08	3.20	3.29	3.36	3.42	3.47	3.52
16	2.58	2.88	3.05	3.17	3.26	3.33	3.39	3.44	3.48
17	2.57	2.86	3.03	3.14	3.23	3.30	3.36	3.41	3.45
18	2.55	2.84	3.01	3.12	3.21	3.27	3.33	3.38	3.42
19	2.54	2.83	2.99	3.10	3.18	3.25	3.31	3.36	3.40
20	2.53	2.81	2.97	3.08	3.17	3.23	3.29	3.34	3.38
24	2.49	2.77	2.92	3.03	3.11	3.17	3.22	3.27	3.31
30	2.46	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.24
40	2.42	2.68	2.82	2.92	2.99	3.05	3.10	3.14	3.18
60	2.39	2.64	2.78	2.87	2.94	3.00	3.04	3.08	3.12
120	2.36	2.60	2.73	2.82	2.89	2.94	2.99	3.03	3.06
inf.	2.33	2.56	2.68	2.77	2.84	2.89	2.93	2.97	3.00

* Table 1b gives a solution $d_i' = t$ to equation (4) in the text for $P=.99$ for the case $\rho_{ij}=1/2$.

TABLE 2a*

TABLE OF t FOR TWO-SIDED COMPARISONS BETWEEN p TREATMENT MEANS AND A CONTROL FOR A JOINT CONFIDENCE COEFFICIENT OF $P=95\%$

	p, Number Of Treatment Means (Excluding The Control)								
d.f.	1	2	3	4	5	6	7	8	9
5	2.57	3.03	3.39	3.66	3.88	4.06	4.22	4.36	4.49
6	2.45	2.86	3.18	3.41	3.60	3.75	3.88	4.00	4.11
7	2.36	2.75	3.04	3.24	3.41	3.54	3.66	3.76	3.86
8	2.31	2.67	2.94	3.13	3.28	3.40	3.51	3.60	3.68
9	2.26	2.61	2.86	3.04	3.18	3.29	3.39	3.48	3.55
10	2.23	2.57	2.81	2.97	3.11	3.21	3.31	3.39	3.46
11	2.20	2.53	2.76	2.92	3.05	3.15	3.24	3.31	3.38
12	2.18	2.50	2.72	2.88	3.00	3.10	3.18	3.25	3.32
13	2.16	2.48	2.69	2.84	2.96	3.06	3.14	3.21	3.27
14	2.14	2.46	2.67	2.81	2.93	3.02	3.10	3.17	3.23
15	2.13	2.44	2.64	2.79	2.90	2.99	3.07	3.13	3.19
16	2.12	2.42	2.63	2.77	2.88	2.96	3.04	3.10	3.16
17	2.11	2.41	2.61	2.75	2.85	2.94	3.01	3.08	3.13
18	2.10	2.40	2.59	2.73	2.84	2.92	2.99	3.05	3.11
19	2.09	2.39	2.58	2.72	2.82	2.90	2.97	3.04	3.09
20	2.09	2.38	2.57	2.70	2.81	2.89	2.96	3.02	3.07
24	2.06	2.35	2.53	2.66	2.76	2.84	2.91	2.96	3.01
30	2.04	2.32	2.50	2.62	2.72	2.79	2.86	2.91	2.96
40	2.02	2.29	2.47	2.58	2.67	2.75	2.81	2.86	2.90
60	2.00	2.27	2.43	2.55	2.63	2.70	2.76	2.81	2.85
120	1.98	2.24	2.40	2.51	2.59	2.66	2.71	2.76	2.80
inf.	1.96	2.21	2.37	2.47	2.55	2.62	2.67	2.71	2.75

* Table 2a gives a solution $d_t'' = t$ which makes the right-hand side of inequality (11) in the text equal to .95 for the case $\rho = 1/2$. This may be used as an approximate solution to equation (5) in the text for $P = 95$ for the case $\rho_{ij} = 1/2$.

TABLE 2b*

TABLE OF t FOR TWO-SIDED COMPARISONS BETWEEN p TREATMENT MEANS AND A CONTROL FOR A JOINT CONFIDENCE COEFFICIENT OF $P=99\%$

	p, Number Of Treatment Means (Excluding The Control)								
d.f.	1	2	3	4	5	6	7	8	9
5	4.03	4.63	5.09	5.44	5.73	5.97	6.18	6.36	6.53
6	3.71	4.22	4.60	4.88	5.11	5.30	5.47	5.61	5.74
7	3.50	3.95	4.28	4.52	4.71	4.87	5.01	5.13	5.24
8	3.36	3.77	4.06	4.27	4.44	4.58	4.70	4.81	4.90
9	3.25	3.63	3.90	4.09	4.24	4.37	4.48	4.57	4.65
10	3.17	3.53	3.78	3.95	4.10	4.21	4.31	4.40	4.47
11	3.11	3.45	3.68	3.85	3.98	4.09	4.18	4.26	4.33
12	3.05	3.39	3.61	3.76	3.89	3.99	4.08	4.15	4.22
13	3.01	3.33	3.54	3.69	3.81	3.91	3.99	4.06	4.13
14	2.98	3.29	3.49	3.64	3.75	3.84	3.92	3.99	4.05
15	2.95	3.25	3.45	3.59	3.70	3.79	3.86	3.93	3.99
16	2.92	3.22	3.41	3.55	3.65	3.74	3.82	3.88	3.93
17	2.90	3.19	3.38	3.51	3.62	3.70	3.77	3.83	3.89
18	2.88	3.17	3.35	3.48	3.58	3.67	3.74	3.80	3.85
19	2.86	3.15	3.33	3.46	3.55	3.64	3.70	3.76	3.81
20	2.85	3.13	3.31	3.43	3.53	3.61	3.67	3.73	3.78
24	2.80	3.07	3.24	3.36	3.45	3.52	3.58	3.64	3.69
30	2.75	3.01	3.17	3.28	3.37	3.44	3.50	3.55	3.59
40	2.70	2.95	3.10	3.21	3.29	3.36	3.41	3.46	3.50
60	2.66	2.90	3.04	3.14	3.22	3.28	3.33	3.38	3.42
120	2.62	2.84	2.98	3.08	3.15	3.21	3.25	3.30	3.33
inf.	2.58	2.79	2.92	3.01	3.08	3.14	3.18	3.22	3.25

* Table 2b gives a solution $d_i''=t$ which makes the right-hand side of inequality (11) in the text equal to .99 for the case $\rho=1/2$. This may be used as an approximate solution to equation (5) in the text for $P=.99$ for the case $\rho_{ij}=1/2$

REFERENCES

[1] Bechhofer, Robert E., "A single-sample multiple decision procedure for ranking means of normal populations with known variances," *Annals of Mathematical Statistics*, 25 (1954), 16–39.

[2] Bechhofer, Robert E., Dunnett, Charles W., and Sobel, Milton, "A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance," *Biometrika*, 41 (1954), 170–6.

[3] Bliss, C. I., *The Statistics of Bioassay*. New York, Academic Press, 1952

[4] Dunnett, Charles W., and Sobel, Milton, "A bivariate generalization of Student's *t*-distribution, with tables for certain special cases," *Biometrika*, 41 (1954), 153–69.

[5] Dunnett, C. W., and Sobel, M., "Approximations to the probability integral and certain percentage points of a multivariate analogue of Student's *t*-distribution," *Biometrika*, 42 (1955), 258–60.

[6] Fieller, E. C., "A fundamental formula in the statistics of biological assay, and some applications," *Quarterly Journal of Pharmacy and Pharmacology*, 17 (1944), 117–23.

[7] Finney, D. J., *Statistical Method in Biological Assay* New York: Hafner Publishing Company, 1952.

[8] Hartley, H. O, and Pearson, E S., "Table of the probability integral of the *t*-distribution," *Biometrika*, 37 (1950), 168–72

[9] National Bureau of Standards, *Tables of Normal Probability Functions*. Washington: U. S. Government Printing Office, 1953

[10] National Bureau of Standards, Unpublished tables

[11] Paulson, Edward, "On the comparison of several experimental categories with a control," *Annals of Mathematical Statistics*, 23 (1952), 239–46.

[12] Pearson, E. S., and Hartley, H. O., *Biometrika Tables for Statisticians, Vol. I.* Cambridge: Cambridge University Press, 1954.

[13] Pearson, K., *Tables for Statisticians and Biometricians*, Part II. Cambridge: Cambridge University Press, 1931

[14] Pillai, K. C. S., and Ramachandran, K V, "On the distribution of the ratio of the *i*th observation in an ordered sample from a normal population to an independent estimate of the standard deviation," *Annals of Mathematical Statistics*, 25 (1954), 565–72

[15] Roessler, E B., "Testing the significance of observations compared with a control," *Proceedings of the American Society for Horticultural Science*, 47 (1946), 249–51.

[16] Scheffé, Henry, "A method for judging all contrasts in the analysis of variance," *Biometrika*, 40 (1953), 87–104

[17] Tukey, John W., "The problem of multiple comparisons" Unpublished notes in private circulation Princeton University.

[18] Villars, Donald Statler, *Statistical Design and Analysis of Experiments for Development Research*. Dubuque, Iowa: Wm. C. Brown Company, 1951.

MAXIMUM LIKELIHOOD ESTIMATION OF THE DISPERSION PARAMETER OF A CHI-DISTRIBUTED RADIAL ERROR FROM TRUNCATED AND CENSORED SAMPLES WITH APPLICATIONS TO TARGET ANALYSIS*

A. CLIFFORD COHEN, JR.
The University of Georgia

The radial error, $r=(x_1^2+x_2^2+\cdots+x_p^2)^{1/2}$, where components $x_j(j=1, 2, \cdots p)$ constitute a set of p independent random variables, each of which is normal $(0, \sigma)$, has the frequency function $[2r/\sigma^2]f_p(r^2/\sigma^2)$ in which $f_p(\chi^2)$ is the χ^2 frequency function with p degrees of freedom. This distribution is of interest in a class of target analysis problems where the origin or "center of impact" is known or can be assumed as known. For $p=2$ and $p=3$, σ is important as a measure of error dispersion and hence as a measure of weapon-system accuracy. Here, we are concerned with estimating σ from truncated and censored samples as well as from complete samples. Maximum likelihood estimating equations are obtained in these cases, and where explicit estimators are not possible, the equations are reduced to forms which permit solution by interpolation with the aid of normal curve tables of areas and ordinates. To facilitate solution in the cases of truncated samples for $p=2$ and $p=3$, tables and graphs of the estimating functions are given. An explicit estimator is obtained in the case of a censored sample when $p=2$. Asymptotic variances of the estimates are obtained for each of the cases considered. Illustrative examples relating to target analysis studies are included. For use when the origin is not known, a graphical procedure for estimating its coordinates based on use of an "estimating circle" of fixed radius, located to enclose a maximum number of sample points or "hits" is described for the cases in which $p=2$.

I. INTRODUCTION

This paper deals specifically with the distribution of a generalized p-dimensional radial error defined as

$$r = \sqrt{x_1^2 + x_2^2 + \cdots + x_p^2}, \tag{1}$$

where components x_j $(j=1, 2, \cdots p)$ are independent random variables, each of which is normally distributed about a known mean with standard deviation σ. It is well known (c.f., for example, Cramér [3],

* Sponsored by the Office of Ordnance Research, U. S. Army under contract DA-01-009-ORD-288

p. 233) that $(r/\sigma)^2$ has a chi-square distribution with p degrees of freedom; that is,

$$f_p(r^2/\sigma^2) = \frac{2^{-p/2}}{\Gamma(p/2)}\left(\frac{r}{\sigma}\right)^{2(p/2-1)} \exp\left[-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right]. \qquad (2)$$

It follows that (r/σ) has a chi distribution, and for the frequency function of r, which is of primary interest here, we have

$$f_p(r) = \frac{2^{-(p-2)/2}}{\sigma\Gamma(p/2)}\left(\frac{r}{\sigma}\right)^{p-1} \exp\left[-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right], \quad 0 \leqq r \leqq \infty \qquad (3)$$

With reference to $f_p(r)$, σ is to be regarded as a parameter of scale or of dispersion. The frequency function $f_p(r)$ is a generalization of the two dimensional frequency function

$$f_2(r) = (r/\sigma^2) \exp\left[-(r/\sigma)^2/2\right], \qquad (4)$$

of radial errors on a plane target in which errors are distances from the target center (or a known center of impact) to points where missiles strike the target

The dispersion parameter, σ, of (4) is of interest in numerous target analysis problems as a measure of weapon-system accuracy, and it is often of practical importance to estimate it, not only from complete samples, but from samples that may be truncated or censored. To be complete, all sample values of r must be measured When rounds miss a target, and their errors thus cannot be measured, truncated or censored samples result. A truncated sample is one in which all r exceeding a known terminal, r_0, are omitted, and the number of such omissions is unknown, the effect being the same as if all errors for which $r > r_0$ were eliminated from the population. A censored sample is one in which the number of errors exceeding r_0 is known, but such errors are not otherwise measured. In both truncated and censored samples, actual measurements are recorded for all $r \leqq r_0$

The work reported on here is related to previous investigations of truncated samples from Type III distributions by the writer [2] and by Des Raj [4]. It is also related to the estimation of parameters of univariate normal populations, a subject which has been considered by numerous writers (c.f., [1] and other references listed therein).

II. COMPLETE SAMPLES

Consider random selections from a population distributed according to (3) when observation of r is unrestricted so that any value within the possible range $0 \leqq r \leqq \infty$ can be observed and measured A sample con-

sisting of n such observations is said to be complete (or unrestricted) of size n, and its likelihood function is given by

$$P = 2^{-n(p-2)/2}\sigma^{-pn}[\Gamma(p/2)]^{-n}\left[\prod_1^n r_i^{p-1}\right]\exp\left[-\sum_1^n r_i^2/2\sigma^2\right]. \tag{5}$$

Taking logarithms, differentiating and equating to zero, we have

$$\frac{\partial L}{\partial\sigma} = -\frac{pn}{\sigma} + \frac{1}{\sigma^3}\sum_1^n r_i^2 = 0, \tag{6}$$

where $L = \log P$. The maximum likelihood estimator of σ follows as

$$\hat{\sigma} = \sqrt{\sum_1^n r_i^2/pn}, \tag{7}$$

where the symbol (^) serves to distinguish the estimator from the parameter estimated, a convention observed throughout this paper.

III. TRUNCATED SAMPLES—NUMBER OF ELIMINATED OBSERVATIONS UNKNOWN

The likelihood function of a sample consisting of n random observations from a population distributed according to (3) when each observation is subject to the restriction $0 \leq r \leq r_0$ is

$$\begin{aligned} P = {} & 2^{-n(p-2)/2}\sigma^{-pn}[\Gamma(p/2)]^{-n}\left[\prod_1^n r_i^{p-1}\right][1 - I_p]^{-n} \\ & \cdot\exp\left[-\sum_1^n r_i^2/2\sigma^2\right], \end{aligned} \tag{8}$$

where

$$I_p(r_0:\sigma) = \frac{2^{-(p-2)/2}}{\Gamma(p/2)}\int_{r_0}^{\infty}\left\{\left(\frac{r}{\sigma}\right)^{p-1}\exp\left[-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2\right]\right\}\frac{dr}{\sigma}. \tag{9}$$

If we let

$$\xi = r/\sigma \quad \text{and thus} \quad \xi_0 = r_0/\sigma, \tag{10}$$

equation (9) reduces to

$$\begin{aligned} I_p(r_0:\sigma) &= I_p(\xi_0) = \int_{\xi_0}^{\infty}\theta_p(\xi)d\xi, \quad \text{where} \\ \theta_p(\xi) &= \frac{2^{-(p-2)/2}}{\Gamma(p/2)}\xi^{p-1}\exp(-\xi^2/2). \end{aligned} \tag{11}$$

Taking logarithms of (8), differentiating and equating to zero, we have

$$\frac{\partial L}{\partial \sigma} = \frac{n}{1 - I_p}\frac{\partial I_p}{\partial \sigma} - \frac{pn}{\sigma} + \frac{1}{\sigma^3}\sum_1^n r_i^2 = 0.$$

Differentiating (11), we obtain

$$\frac{\partial I_p(\xi_o)}{\partial \sigma} = \frac{\xi_o}{\sigma}\,\theta_p(\xi_o), \tag{12}$$

and with this result, the preceding equation reduces to

$$\frac{\partial L}{\partial \sigma} = \frac{n\xi_o}{\sigma}\,\frac{\theta_p(\xi_o)}{1 - I_p(\xi_o)} - \frac{pn}{\sigma} + \frac{1}{\sigma^3}\sum_1^n r_i^2 = 0. \tag{13}$$

Equation (13) further reduces to

$$\sum_1^n r_i^2/nr_o^2 = G_p(\xi_o), \tag{14}$$

where

$$G_p(\xi_o) = \frac{1}{\xi_o}\left[\frac{p}{\xi_o} - \frac{\theta_p(\xi_o)}{1 - I_p(\xi_o)}\right] \tag{15}$$

Equation (14) can be solved for $\hat{\xi}_o$ using standard iterative procedures or by interpolation in appropriate tables. From (10) it follows that

$$\hat{\sigma} = r_o/\hat{\xi}_o. \tag{16}$$

Truncated Sample From Population of Dimension Two When $p=2$, $\theta_2(\xi_o) = \xi_o \exp - \xi_o^2/2 = \sqrt{2\pi}\xi_o\phi(\xi_o)$, and $I_2(\xi_o) = \sqrt{2\pi}\phi(\xi_o)$, where $\phi(t) = (1/\sqrt{2\pi}) \exp - t^2/2$, the standardized normal curve ordinate. For this case, estimating equation (14) becomes

$$\frac{\sum_1^n r_i^2}{nr_o^2} = \frac{2}{\hat{\xi}_o^2} - \frac{\phi(\hat{\xi}_o)}{\phi(0) - \phi(\hat{\xi}_o)} = G_2(\hat{\xi}_o) \tag{17}$$

For any given ξ, $G_2(\xi)$ can be evaluated from an ordinary table of normal curve ordinates, and simple interpolation can be used to solve (17) for $\hat{\xi}_o$. In view of the particular interest attached to this case, an abbreviated tabulation of $G_2(\xi)$ is included in Table 1, and a graph of this function is given in Fig. 1. With $\sum_1^n r_i^2/nr_o^2$ computed from the sample, $\hat{\xi}_o$ can be read directly from the graph with sufficient accuracy for many purposes. More precise values can be obtained by interpola-

TABLE 1

THE FUNCTIONS $G_2(\xi)$ AND $G_3(\xi)$

$$G_2(\xi)=\frac{2}{\xi^2}-\frac{\phi(\xi)}{\phi(0)-\phi(\xi)},\qquad G_3=\frac{3}{\xi^2}-\frac{2\xi\phi(\xi)}{1-[2\xi\phi(\xi)+I_1(\xi)]}.$$

ξ	$G_2(\xi)$	$G_3(\xi)$	ξ	$G_2(\xi)$	$G_3(\xi)$
0	.50000	.60000	2.0	34348	.45758
1	49958	.59965	2.1	.32960	44330
.2	.49833	59863	2.2	.31562	.42857
3	49625	.59691	2 3	.30164	41349
.4	.49333	59450	2 4	.28775	.39814
.5	.48959	.59138	2.5	.27404	.38261
6	.48501	58756	2 6	.26061	.36702
7	.47960	58303	2.7	24753	.35145
.8	.47338	.57777	2 8	.23486	.33602
9	.46634	57178	2 9	22267	.32082
1 0	.45851	56505	3 0	21099	.30594
1.1	44989	55757	3.1	19986	.29146
1.2	.44051	54935	3 2	18930	.27745
1.3	43041	54037	3 3	17932	.26397
1.4	41961	.53064	3 4	16991	.25106
1.5	40817	52018	3 5	.16107	.23875
1 6	.39614	50899	3 6	.15278	.22705
1 7	.38358	49710	3 7	.14503	.21598
1 8	37056	.48453	3.8	.13777	.20553
1 9	.35717	47134	3 9	13099	19569
			4 0	.12466	.18643
			4 5	09873	.14800
			5 0	.08000	.11999

tion in Table 1, and when still greater accuracy is demanded, additional values of $G_2(\xi)$ can be computed from the normal curve ordinates for a finer interval of interpolation. Once $\hat{\xi}_*$ has been determined with the necessary accuracy, $\hat{\sigma}$ follows from (16).

Truncated Sample From Population of Dimension Three. Radial errors of dimension three are of interest in certain weapon-system accuracy studies involving three dimensional targets The distribution of r in this case is also of interest in studies of molecule velocity with respect to a three dimensional system of rectangular coordinates in which r is interpreted as velocity rather than as target error. In view of this practical interest, more explicit estimators are given.

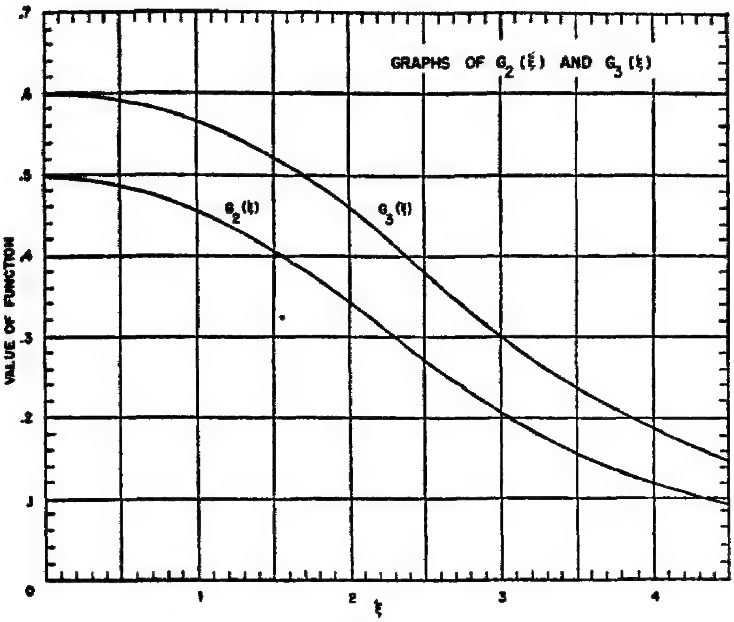


Fig 1

With $p=3$, $\theta_3(\xi_o)=2\xi_o^2\phi(\xi_o)$, $I_3(\xi_o)=[2\xi_o\phi(\xi_o)+I_1(\xi_o)]$, and equation (14) becomes

$$\frac{\sum_1^n r_i^2}{nr_0^2} = \frac{3}{\hat{\xi}_o^2} - \frac{2\hat{\xi}_o\phi\hat{\xi}(_o)}{1 - [2\hat{\xi}_o\phi(\hat{\xi}_o) + I_1(\xi_o)]} = G_3(\hat{\xi}_o) \tag{18}$$

We note that $I_1(\xi_o)=2\int_{\xi_0}^\infty\phi(t)dt$, which is double the area under the normal curve to the right of $\xi=\xi_o$. The manner of solving (18) for $\hat{\xi}_o$ is essentially the same as in the two dimensional case To facilitate this solution, $G_3(\xi)$ is also included in Table 1 and its graph is given in Fig 1 along with that of $G_2(\xi)$

IV. CENSORED SAMPLES—NUMBER OF UNMEASURED OBSERVATIONS KNOWN

If we make random observations on a population distributed according to (3), record measurements of each error for which $0 \leq r \leq r_o$, while counting but not otherwise measuring those for which $r > r_o$ until

n measured observations have been accumulated, the sample obtained is said to be censored. Let n_e designate the number of censored observations for which $r > r_0$, and the likelihood function for a sample of this type may be written as

$$P = k2^{-n(p-2)/2}\sigma^{-pn}[\Gamma(p/2)]^{-n}\left[\prod_1^n r_i^{p-1}\right]I_p^{n_e} \cdot \exp\left[-\sum_1^n r_i^2/2\sigma^2\right], \tag{19}$$

where k is a constant and $I_p(\xi_0)$ is given by (11). Taking logarithms of (19), differentiating and equating to zero, we obtain

$$\frac{\partial L}{\partial \sigma} = \frac{n_e}{I_p}\frac{\partial I_p}{\partial \sigma} - \frac{pn}{\sigma} + \frac{1}{\sigma^3}\sum_1^n r_i^2 = 0. \tag{20}$$

Using (12), this result reduces to

$$\frac{\sum_1^n r_i^2}{nr_0^2} = \frac{1}{\hat{\xi}_0}\left[\frac{p}{\hat{\xi}_0} - \frac{n_e}{n}\frac{\theta_p(\hat{\xi}_0)}{I_p(\hat{\xi}_0)}\right] = H_p(\hat{\xi}_0). \tag{21}$$

If instead of fixing n, the number of measured observations, in advance of selecting a sample, we fix the total number of observations, $n+n_e$, the same likelihood function (19) and the same estimating equation (21) are obtained. In this case, the estimating equation would not have a solution when $n=0$, but this limitation is of little practical importance. The solution of (21) for $\hat{\xi}_0$ in the general p-dimensional case is essentially the same as for (14) in the truncated case.

Samples considered here are of the fixed terminal type; that is, r_0 is fixed prior to sampling. In some situations, fixed size (variable terminal) censored samples may arise in which the n_e (fixed) largest errors out of a fixed total number of observations are unmeasured. Estimating equations for samples of this latter type are, however, identical with those given here for corresponding fixed-terminal samples if r_0 is replaced with r_n, the largest of the measured sample observations.

Censored Sample From Population of Dimension Two. When $p=2$, $\theta_2(\xi_0)/I_2(\xi_0)=\xi_0$, and (21) becomes

$$\frac{\sum_1^n r_i^2}{nr_0^2} = \frac{2}{\hat{\xi}_0^2} - \frac{n_e}{n}. \tag{22}$$

Using (10), this reduces to

$$\hat{\sigma} = \sqrt{\left[\sum_1^n r_i^2 + n_o r_o^2\right]/2n}, \tag{23}$$

which, in this case, is a simple explicit estimator of the required parameter.

Censored Sample From Population of Dimension Three. When $p=3$, $\theta_3(\xi_o)/I_3(\xi_o)=\xi_o^2/[\xi_o+I_1(\xi_o)/2\phi(\xi_0)]$, where as previously indicated $I_1(\xi_o)$ is twice the area under the normal curve to the right of $\xi=\xi_o$ and $\phi(\xi_o)$ is the normal curve ordinate. In this case, (21) becomes

$$\frac{\sum_1^n r_i^2}{nr_o^2} = \frac{3}{\hat{\xi}_o^2} - \frac{n_o\hat{\xi}_o}{n[\hat{\xi}_o + I_1(\hat{\xi}_o)/2\phi(\hat{\xi}_o)]}. \tag{24}$$

With the aid of tables of normal curve areas and ordinates, this equation can be solved for $\hat{\xi}_o$ by interpolation as already indicated for the general case of (21). For a first approximation, the value of ξ read from the G_3 curve of Fig. 1 should suffice.

V. RELIABILITY OF ESTIMATES

Complete Samples. For complete samples, $(pn\hat{\sigma}^2/\sigma^2)$ has a chi-square distribution with pn degrees of freedom To prove this, we note that each component x_j of r is normal $(0, \sigma)$ and thus x_j/σ is normal $(0, 1)$. Let x_{ji} designate the ith observation of x_j, and $\sum_1^n (x_{ji}/\sigma)^2$, which is $(x_j/\sigma)^2$ summed over all sample observations, has a chi-square distribution with d.f.$=n$ Since the sum of independent chi-square distributions is itself a chi-square distribution with degrees of freedom equal to the sum of the respective degrees of freedom of the component distributions, and since $\sum_{j=1}^p \sum_{i=1}^n (x_{ji}/\sigma)^2=(pn\hat{\sigma}^2/\sigma^2)$, the stated result follows.

Confidence intervals for σ can be determined as

$$P\left\{\sqrt{\frac{pn\hat{\sigma}^2}{\chi_2^2}} < \sigma < \sqrt{\frac{pn\hat{\sigma}^2}{\chi_1^2}}\right\} = 1 - \alpha, \tag{25}$$

where $\hat{\sigma}$ is given by (7) and χ_1^2 and χ_2^2 are read from standard chi-square tables with d.f.$=pn$ such that $P[\chi^2>\chi_1^2]=1-\alpha/2$, and $P[\chi^2>\chi_2^2]=\alpha/2$.

Since $(pn\hat{\sigma}^2/\sigma^2)$ has a chi-square distribution with d.f.$=pn$, then $(\sqrt{pn}/\sigma)\hat{\sigma}$ has a chi distribution with the same degrees of freedom.

Kendall [5], p. 294 gives the moments of the chi distribution, and using his results, the exact variance of $\hat{\sigma}$ is

$$V(\hat{\sigma}) = \frac{\sigma^2}{2pn}\left[2pn - \left\{\frac{2\Gamma\left(\frac{pn+1}{2}\right)}{\Gamma\left(\frac{pn}{2}\right)}\right\}^2\right]. \tag{26}$$

An expansion based on an extended form of Stirling's formula, also given by Kendall permits (26) to be written as

$$V(\hat{\sigma}) = \frac{\sigma^2}{2pn}\left[1 - \frac{1}{4pn} + \cdots\right]. \tag{27}$$

For the mean, Kendall's results give

$$m_{\hat{\sigma}} = \sigma\left[1 - \frac{1}{4pn} + \frac{1}{32(pn)^2} + \cdots\right] \tag{28}$$

For large values of n, confidence intervals for σ can be approximated from the normal distribution with mean and variance as given by (27) and (28).

Similar exact sampling results are not available for restricted samples, but asymptotic variances for each of the maximum likelihood estimates, including that based on a complete sample, are given by

$$\text{Asy Var.}(\hat{\sigma}) = -\left[E\left(\frac{\partial^2 L}{\partial\sigma^2}\right)\right]^{-1}. \tag{29}$$

For restricted samples, this leads to specific variances as given below

Truncated Samples. Although the result in this case is not claimed to be obvious, straightforward manipulation gives

$$E\left(\frac{\partial^2 L}{\partial\sigma^2}\right) = -\frac{2pE(n)}{\sigma^2}\left\{1 - \frac{\xi_o\theta_p(\xi_o)}{2p[1 - I_p(\xi_o)]}\left[\xi_o^2 - (p-2) + \frac{\xi_o\theta_p(\xi_o)}{[1 - I_p(\xi_o)]}\right]\right\} \tag{30}$$

Using this result, the asymptotic variance in the *two-dimensional case* is given by (29) as

$$\text{Asy. Var.}(\hat{\sigma}) = \frac{\sigma^2}{4E(n)}\left\{1 - \frac{\xi_o^4}{4}\left(\frac{\phi(0)\phi(\xi_o)}{[\phi(0) - \phi(\xi_o)]^2}\right)\right\}^{-1}. \tag{31}$$

Censored Samples. Again in this case, after some straightforward though involved manipulation, we find

$$E\left(\frac{\partial^2 L}{\partial \sigma^2}\right) = -\frac{2pE(n)}{\sigma^2}\left\{1 - \frac{\xi_o \phi_p(\xi_o)}{2p[1 - I_p(\xi_o)]}\left[\xi_o^2 - (p-2) - \frac{\xi_o \theta_p(\xi_o)}{I_p(\xi_o)}\right]\right\}. \tag{32}$$

Using this result, the corresponding asymptotic variance for a *two-dimensional population* is given by (29) as

$$\text{Asy. Var. } (\hat{\sigma}) = \sigma^2/4E(n). \tag{33}$$

Throughout this paper, n designates the number of measured observations in each sample considered, and $E(n)$ which appears in equations (30–33) is the expected value of n. With $E(n)$ appropriately evaluated, the same variance formulas are applicable regardless of the sampling scheme employed Values of $E(n)$ for three sampling schemes that seem most likely to occur are given below.

(i) The experiment is continued until n measured observations are obtained in the range $0 \leqq r \leqq r_o$ Since n is a fixed number in this case, $E(n) = n$

(ii) A fixed total of say, N, $(=n+n_o)$ observations are selected, but only those for which $0 \leqq r \leqq r_o$ are measured The number of measured observations, n, in this case is a random variable and $E(n) = N[1 - I_p(\xi_o)]$

(iii) The n smallest observations out of a total of N are measured. In this case both n and N are fixed numbers, but r_o is a random variable. As already mentioned the estimators in this case are the same as those given for the censored samples of Section IV, provided r_o is replaced by r_n, where r_n is the largest of the n measured observations in a given sample. With n fixed, $E(n) = n$ as in case (i) above

VI ILLUSTRATIVE EXAMPLES

To illustrate the practical application of results obtained in this paper, we employ a sample selected from a population distributed according to (3) with $p=2$ and $\sigma=10$. The actual selection was made using a table of random numbers, and although certain approximations were involved in obtaining observations, the sample appears to be adequate for the intended purpose. By appropriately adding information, the same basic data of the truncated sample also serves to illustrate estimation from censored and complete samples.

Truncated Sample. In this case, we let $r_c = 26$, and the sample consists of $n = 25$ observations, each of which is less than or equal to r_c. $\sum_1^{25} r_i^2 = 3561$, and $\sum_1^{25} r_i^2/nr_c^2 = 3561/(25)(26)^2 = 0.21071$. Interpolating from $G_2(\xi)$ in Table 1, we have $\hat{\xi}_0 = 3.003$, a result which might be read with only slightly less accuracy from the corresponding graph of Fig. 1. From (16), we have $\hat{\sigma} = 26/3.003 = 8.66$. Using this value for σ, which for illustrative purposes, is assumed to be unknown, we obtain $V(\hat{\sigma})$ from (31) as 1.27481 and $\sigma_{\hat{\sigma}} = \sqrt{V(\hat{\sigma})} = 1\ 129$.

Censored Sample. To the truncated sample, we add information of the occurrence of $n_e = 2$ observations for which $r > r_c$. Otherwise, the sample remains unchanged. The required estimate is computed from (23) as $\hat{\sigma} = \sqrt{[3561 + 2(26)^2]/2(25)} = 9.91$. Using this value for σ which is still assumed to be unknown (for purposes of this illustration), we approximate $V(\hat{\sigma})$ from (33) as $9.91^2/4(25) = 0.9826$, and $\sigma_{\hat{\sigma}} = 0.991$.

Complete Sample. To the censored sample, actual measurements of the two censored observations were added to produce a complete sample with $n = 27$, and for the sample thus formed, $\sum_1^{27} r_i^2 = 5019$. From (7) with $p = 2$, we compute $\hat{\sigma} = \sqrt{5019/2(27)} = 9.64$, a value which differs by only a small amount from the censored sample estimate of the previous paragraph. The variance of this estimate is computed from (27) as $V(\hat{\sigma}) = 0.8520$, and $\hat{\sigma}_{\hat{\sigma}} = 0\ 923$.

The above results are summarized in the following table along with 0.95 confidence limits based on the variances as computed and the normal curve approximation for the distribution of $\hat{\sigma}$ in each of the cases considered.

Type of Sample	$\hat{\sigma}$	0.95 Confidence Interval
Truncated	8 66	$6.45 < \sigma < 10.87$
Censored	9 91	$7.97 < \sigma < 11.85$
Complete	9.64	$7.79 < \sigma < 11.41$

VII. SOME FURTHER REMARKS

The origin from which radial errors are measured is not known in all target analyses and its coordinates which are means of the component distributions must be estimated from the sample. This situation presents no essential difficulties in unrestricted samples since sample means are maximum likelihood estimators of population means, but in restricted samples it introduces some major complications. In the two-dimensional case, however, an estimating circle to be located graphically can be employed to advantage. Consider the two-dimensional

plane of the distribution of x and y. Require that a circle of fixed radius be so located that the probability of a pair (x, y) falling in that circle is a maximum. The center of such a circle must be located at the point (m_x, m_y). Since m_x and m_y are unknown, we locate our "estimating circle" by placing it over the scatter diagram of an observed sample and shifting its position until the maximum number of points or "hits" is enclosed. The coordinates of the center of this circle are then taken as estimates of the means, m_x and m_y. With small samples, we may not always find a unique position for a circle of fixed radius and it may be necessary to adjust the radius to alleviate this difficulty. A circle which encloses from 70–90 per cent of all sample points will be satisfactory in ordinary circumstances.

As an alternate method of determining estimates, we may locate the estimating circle as described above and based only on points inside this circle, use as estimators

$$m_x^{**} = \sum_1^{n} x_i/n_k, \quad \text{and} \quad m_y^{**} = \sum_1^{n_k} y_i/n_k,$$

where n_k is the number of points inside the circle.

When means are estimated as suggested in this section, the variances of Section V are not strictly applicable. If, however, the number of points neglected in determining estimates of the means is small, these variances should furnish reasonable approximations to the correct values.

An example of hits of a hand thrown dart against a rectangular plane target affords an opportunity to illustrate how estimates of this section are determined. Out of 20 throws at the target, one resulted in a miss The scatter diagram of the remaining 19 hits is shown to scale in Fig 2. The coordinates of these 19 hits referred to the lower left corner of the target as an origin are given below.

x	y	x	y	x	y	x	y	x	y
1.3	4.9	2.6	5.2	3.2	5 6	5.3	6 8	5.0	3.0
1.4	5.5	2.7	6 3	3.7	3.6	5 6	5 4	6.5	6.2
1.8	3.8	2 8	6.7	3 9	5 2	5 7	5 0	1.7	2.6
2.3	3.7	3.6	6 8	4.8	7.0	5.0	3 6		

An estimating circle of radius r_e was located so that it encloses 15 points, which is the largest number that can be enclosed without increasing the radius. Coordinates of the center of this circle (3.5, 5.2)

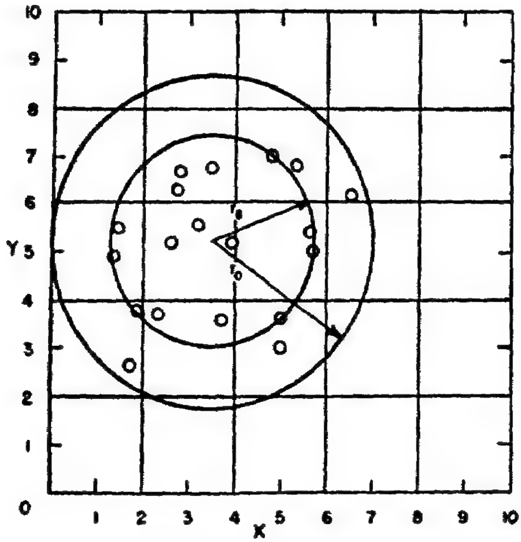


FIG 2 Scatter diagram of darts on a target.

are estimates of m_x and m_y respectively The terminal radius is now the distance from the center of the estimating circle to the nearest edge of the target For our sample, $r_o=3\ 50$, and we consider any hit outside the circle thus defined as a censored observation. If such observations are treated otherwise, estimator (23) will not be applicable. Although this results in some loss of information in certain cases, the loss is almost negligible if r_o exceeds two standard deviations. In the present illustration, only the single throw which missed the target entirely gave a point outside this circle Accordingly, we now have a censored sample for which $n=19$, $n_o=1$, and $r_o=3.5$ Using the center of the estimating circle as the origin, we compute $\sum_1^{19} r_i^2=80\ 91$ for the 19 observed hits, and from (23) we estimate σ as

$$\sigma^{**} = \sqrt{[80\ 91 + 1(3\ 5)^2]/2(19)} = 1.57.$$

The designation (**) serves to recognize that the estimate is not to be considered as a maximum likelihood estimate in view of the manner of estimating m_x and m_y The radius of the estimating circle in this instance is 1.43 times the standard deviation, and in standard units, the terminal radius, $\xi_o=3\ 5/1\ 57=2\ 23$.

REFERENCES

[1] Cohen, A. C., Jr., "Estimating the mean and variance of normal populations from singly truncated and doubly truncated samples," *Annals of Mathematical Statistics*, 21 (1950), 557–69.

[2] Cohen, A. C, Jr., "Estimating parameters of Pearson Type III populations from truncated samples," *Journal of the American Statistical Association*, 45 (1950), 411–23.

[3] Cramér, H, *Mathematical Methods of Statistics* Princeton: Princeton University Press (1946)

[4] Des Raj, "Estimation of the parameters of Type III populations from truncated samples," *Journal of the American Statistical Association*, 48 (1953), 336–49

[5] Kendall, M G, *The Advanced Theory of Statistics Vol 1.* London· Charles Griffin and Co, Ltd, 4th edition (1948).

THE DISTRIBUTION OF THE QUOTIENT OF MAXIMUM VALUES IN SAMPLES FROM A RECTANGULAR DISTRIBUTION

V. N. MURTY

Central Tobacco Research Institute, India, University of North Carolina

When there is a priori knowledge that two samples have been drawn from rectangular populations with the same lower bounds, the hypothesis that both samples have been drawn from the same population can be tested by means of the quotient of the maximum values. The distribution of this statistic is derived, and its properties studied. Explicit expressions for the power function of the test are given, and the table of 5% values of the quotient is given for sample sizes up to ten A numerical example is given

I INTRODUCTION

RIDER [1] gave the distribution of the ratio of the ranges of two independent random samples from a continuous rectangular population

$$f(x) = \begin{cases} \dfrac{1}{\theta_0} & \text{for} \quad 0 \leqq x \leqq \theta_0 \\ 0 & \text{elsewhere} \end{cases} \tag{1}$$

and used it to test the hypothesis that two samples came from the same rectangular population. The distribution which Rider derived would be used to test the hypothesis that two samples came from rectangular populations having the same range, θ_0, irrespective of their starting points, when we have a priori knowledge that the starting point of the populations is the same, the test proposed in this paper appears to be more powerful than the one proposed by Rider.

II. DISTRIBUTION OF THE QUOTIENT OF MAXIMUM VALUES

If L denotes the maximum observation in a random sample of size n from population (1) then the distribution of L is

$$n\theta_0^{-n}L^{n-1}dL. \tag{2}$$

To derive the distribution of the quotient of the maximum values of two independent random samples from population (1), we let L_1 be the maximum value of a sample of size m, and L_2 the maximum value of

a sample of size n. The joint distribution of L_1 and L_2 is then

$$mn\theta_0^{-(m+n)}L_1^{m-1}L_2^{n-1}dL_1dL_2. \tag{3}$$

From this we wish to determine the distribution of

$$u = L_1/L_2.$$

It is convenient to regard L_2 as the abscissa, and L_1 as the ordinate in rectangular co-ordinates and to transform to polar co-ordinates by setting

$$L_1 = r \sin \theta; \qquad L_2 = r \cos \theta.$$

Then (3) becomes a function of r and θ, which can readily be reduced to a function of r and u, where

$$u = \tan \theta.$$

The variable r can be integrated out in the usual manner, giving the distribution of u, as follows:

$$\frac{mn}{m+n} u^{m-1}du \qquad 0 \leqq u \leqq 1 \tag{4}$$

$$\frac{mn}{m+n} u^{-n-1}du \qquad 1 \leqq u \leqq \infty. \tag{5}$$

It will be noted that the distribution is entirely independent of θ_0

III. DESCRIPTION OF THE DISTRIBUTION CURVE

The distribution curve obviously has the positive u-axis as an asymptote. If $m=2$ the left-hand part of the curve is a straight line passing through the origin and has the slope $2n/(n+2)$. The slope has the value 1 when $n=2$ and approaches 2 as $n\to\infty$ If $m>2$ the curve has a zero slope at the origin.

IV MOMENTS

The kth moment about the origin is

$$\mu_k' = \frac{mn}{(m+k)(n-k)}. \tag{6}$$

In particular the mean is

$$\mu_1' = \frac{mn}{(m+1)(n-1)}. \tag{7}$$

It follows that the mean will be less than, equal to, or greater than 1, according as $n-m$ is greater than, equal to, or less than 1.

If $m=n$, (6) becomes

$$\mu_k' = \frac{n^2}{n^2 - k^2}. \tag{8}$$

If μ_i denotes the ith moment about the mean of the distribution, we have after some simplification

$$\begin{aligned}
\mu_2 &= \frac{n^2(2n^2+1)}{(n^2-1)^2(n^2-4)} \\
\mu_3 &= \frac{2n^2(15n^4+7n^2+2)}{(n^2-1)^3(n^2-4)(n^2-9)} \\
\mu_4 &= \frac{3n^2(8n^8+212\,n^6+95n^4+33n^2+12)}{(n^2-1)^4(n^2-4)(n^2-9)(n^2-16)} \\
\alpha_3 &= \frac{2(15n^4+7n^2+2)(n^2-4)^{1/2}}{n(2n^2+1)^{3/2}(n^2-9)} \\
\alpha_4 &= \frac{3(8n^8+212n^6+95n^4+33n^2+12)(n^2-4)}{n^2(2n^2+1)^2(n^2-9)(n^2-16)}
\end{aligned} \tag{9}$$

where $\alpha_3=\mu_3/\mu_2^{3/2}$ and $\alpha_4=\mu_4/\mu_2^2$.
It is easy to verify that

$$\alpha_3 \to 0 \quad \text{and} \quad \alpha_4 \to 6 \quad \text{as} \quad n \to \infty.$$

3. CUMULATIVE DISTRIBUTION FUNCTION

If the probability of the inequality $u \geqq U$ is denoted by $P\{u \geqq U\}$, we have

$$P\{u \geqq U\} = \frac{m}{m+n} U^{-n} \qquad \text{if} \quad U \geqq 1 \tag{10}$$

$$= 1 - \frac{n}{m+n} U^{m} \quad \text{if} \quad U \leqq 1. \tag{11}$$

Formulas (10) and (11) permit quick computation of the probability of getting a result as significant as the one observed. Table 1 has been computed from these formulas.

TABLE 1

5% POINTS OF u FOR SAMPLE SIZES $\leqq 10$

n \ m*	2	3	4	5	6	7	8	9	10
2	3 162	3 464	3 651	3 780	3 874	3 945	3 999	4 046	4 083
3	2 000	2 153	2 252	2 320	2 371	2 410	2 441	2 466	2 487
4	1 607	1 710	1 778	1 826	1 862	1 888	1.911	1.930	1.944
5	1 417	1 496	1 548	1 585	1 613	1 635	1.652	1 669	1 677
6	1 308	1 371	1 415	1 445	1 479	1 486	1 501	1 513	1 524
7	1 239	1 291	1 328	1 354	1 374	1 390	1 403	1 414	1 422
8	1 189	1 236	1 268	1 291	1 308	1 318	1 334	1 343	1 352
9	1 154	1 196	1 224	1 245	1 260	1 274	1 283	1 291	1 299
10	1 128	1 165	1 191	1 219	1 223	1 236	1 245	1 252	1 259

* m denotes the size of the sample having the greater maximum value.

If the critical region of the test is given by

$$u \geqq u_\alpha$$

where α denotes the level of significance then the above table gives the values of u_α for $\alpha = 0\ 05$ If the critical region of the test is given by

$$u \geqq u_1 \quad \text{or}$$

$$u \leqq u_2$$

then we determine u_1 and u_2 from the relation

$$P\{u \geqq u_1\} + P\{u \leqq u_2\} = \text{Size of the test,}$$

which can be rewritten as

$$P\{u \geqq u_1\} + P\left\{\frac{1}{u} \geqq \frac{1}{u_2}\right\} = \text{Size of the test,}$$

and observing that the distribution $1/u$ is the same as that of u with m and n interchanged we can obtain the value of $1/u_2$ and hence u_2 from the above table It should be remembered that the size of the test with this critical region is 2α and not α

VI NUMERICAL EXAMPLE

The following two samples have been drawn at random We wish to test the hypothesis, that they come from the same rectangular population starting at zero

Sample 1.	0 73,	0.92,	0 07,	0 95	
Sample 2	0 92,	0.19,	0.19,	0.29,	0.38,
	0 16,	0.15,	0.84		

We have

$$L_1 = 0.95, \qquad m = 4$$

$$L_1 = 0.92, \qquad n = 8$$

$$\therefore \; u = L_1/L_2 = 1.03$$

The 5% value of u from Table 1 with $m=4$, $n=8$ is

$$u_{.05} = 1.268.$$

We would consequently accept the hypothesis that the two samples came from the same rectangular population.

VII. POWER OF THE TEST

Let the sample of size m come from the rectangular population

$$f(x) = \begin{cases} \dfrac{1}{k\theta_0} & \text{for} \quad 0 \leqq x \leqq k\theta_0 \\ 0 & \text{elsewhere, and } k > 1 \end{cases} \tag{12}$$

and the sample of size n from the population

$$f(x) = \begin{cases} \dfrac{1}{\theta_0} & \text{for} \quad 0 \leqq x \leqq \theta_0 \\ 0 & \text{elsewhere} \end{cases} \tag{13}$$

Then the distribution of u is easily seen to be

$$\frac{mn}{m+n} k^{-m} u^{m-1} du \qquad 0 \leqq u \leqq k \tag{14}$$

$$\frac{mn}{m+n} k^{n} u^{-n-1} du \qquad k \leqq u \leqq \infty. \tag{15}$$

The most powerful test is obtained by using the likelihood ratio principle and in this case we have the test as:

Region of rejection: $u \geqq U_\alpha$

Region of acceptance: $u < U_\alpha$

where U_α is given by

$$\frac{m}{m+n} U_\alpha^{-n} = \alpha \tag{16}$$

Hence the power of the test is

$$P\{u \geqq U_\alpha\} = \frac{m}{m+n}\left(\frac{k}{U_\alpha}\right)^n \qquad \text{if} \quad k < U_\alpha$$
$$= 1 - \frac{n}{m+n}\left(\frac{U_\alpha}{k}\right)^m \quad \text{if} \quad k > U_\alpha \tag{17}$$

If $o<k<1$, the most powerful test is given by

Region of rejection: $u \leqq U_\alpha$

Region of acceptance: $u > U_\alpha$

Hence the power of the test is

$$P\{u \leqq U_\alpha\} = 1 - \frac{m}{m+n}\left(\frac{k}{U_\alpha}\right)^n \quad \text{if} \quad U_\alpha > k$$
$$= \frac{n}{m+n}\left(\frac{U_\alpha}{k}\right)^m \qquad \text{if} \quad U_\alpha < k \tag{18}$$

The test procedure and the corresponding distribution problem remain unaltered except for the slight change indicated below even when the population is given by

$$f(x) = \frac{1}{\theta_0} \quad \text{for} \quad C \leqq x \leqq C + \theta_0$$
$$= 0 \quad \text{elsewhere}$$

where C is any arbitrary constant

The known lower bound would have to be subtracted from each maximum value before the quotient is computed.

REFERENCE

[1] Rider, Paul R., "The distribution of the quotient of ranges in samples from a rectangular population," *Journal of the American Statistical Association*, 46 (1951), 502–7

THE DISTRIBUTION OF THE PRODUCT OF MAXIMUM VALUES IN SAMPLES FROM A RECTANGULAR DISTRIBUTION

PAUL R. RIDER
Wright-Patterson Air Force Base

I. INTRODUCTION

It is the purpose of this note to supplement the paper by Murty [1] by deriving the distribution of the product of maximum values in random samples from the rectangular distribution

$$f(x) = 1, \qquad 0 \leqq x \leqq 1. \tag{1}$$

In order to keep formulas simpler, a unit range has been chosen. For while Murty's range θ_0 drops out in the distribution of the quotient of maximum values, such is not the case for the distribution of the product.

II. DISTRIBUTION OF THE PRODUCT OF MAXIMUM VALUES

If L is the maximum value of a sample of n from the population (1), then, as Murty has shown, the distribution of L is

$$nL^{n-1}dL \tag{2}$$

Let L_1 be the maximum value of a sample of size m and L_2 the maximum value of a sample of size n, the samples being random and independent. The joint distribution of L_1 and L_2 is

$$mnL_1^{m-1}L_2^{n-1}dL_1dL_2. \tag{3}$$

From this we wish to determine the distribution of the product $v = L_1L_2$.

This is simply a matter of replacing, in (3), L_2 by $L_1^{-1}v$ and dL_2 by $L_1^{-1}dv$, and then integrating with respect to L_1 between the limits v and 1, since these values are the minimum and the maximum which L_1 can attain This leads to

$$mn(m - n)^{-1}v^{n-1}(1 - v^{m-n})dv, \qquad m \neq n; \tag{4}$$

$$n^2v^{n-1} \log (1/v)dv, \qquad m = n. \tag{5}$$

III. PRODUCT OF k MAXIMUM VALUES

We shall now consider the product of k maximum values, limiting the discussion to the case of equal sample sizes.

To derive the distribution of the product of three maximum values in samples from the population (1) we replace the variable v in (5) by y

and multiply the result by (2), obtaining

$$n^2 y^{n-1} \log (1/y) L^{n-1} dy dL. \tag{6}$$

Since y is the product of the maximum values of two independent random samples and L is the maximum value of a random sample which is independent of either of the samples involved in the product y, (6) gives the joint distribution of y and L. If we set $v = Ly$, then v is the product of the maximum values of three independent random samples from (1). We replace, in (6), L by $y^{-1}v$ and dL by $y^{-1}dv$. The distribution of v can now be found by integrating with respect to y from $y = u$ to $y = 1$.

Repeating this process again and again, we find that the distribution of the product v of the maximum values of k independent random samples from the rectangular population (1) is given by

$$[n^k/(k-1)!] v^{n-1} [\log (1/v)]^{k-1} dv, \tag{7}$$

a result which can be established by mathematical induction.

IV. MOMENTS

For the distribution of the product of the maximum values of two independent random samples of sizes m and n respectively ($m \neq n$), the moment of order j about the origin is

$$\mu_j' = mn/(m+j)(n+j). \tag{8}$$

For the distribution of the product of the maximum values of k independent random samples, each of size n, the moment of order j about the origin is

$$\mu_j' = n^k/(n+j)^k \tag{9}$$

Expressions for moments about the origin are not simple and will not be listed However, it is not difficult to show that as $n \to \infty$,

$$\alpha_3 = \mu_3/\sigma^3 \to -\sqrt{2},$$
$$\alpha_4 = \mu_4/\sigma^4 \to 6.$$

REFERENCES

[1] Murty, V. N., "The distribution of the quotient of maximum values in samples from a rectangular distribution," *Journal of the American Statistical Association*, 50 (1955), pp. 1136–1141.

[2] Rider, P R., "The distribution of the product of ranges in samples from a rectangular population," *Journal of the American Statistical Association*, 48 (1953), 546–9.

FIXED, MIXED, AND RANDOM MODELS*

M. B. Wilk† and O. Kempthorne‡
Iowa State College

Some explicit questions are raised regarding the adequacy of assumed linear models as a basis for the interpretation of the analysis of variance of randomized experiments. A generally applicable method for the derivation of a linear statistical model, based on the experimental situation and the design of the experiment, is exemplified. The central features of the method are the notion of "experimental unit," the concept of "true response," and the use of randomization in the design. A model is derived for the case where two factors having A and B levels respectively, are to be examined with respect to a population of P experimental units, where selection of levels of the factors to be tested, selection of experimental units to be used, and the allocation of selected treatment combinations to units is at random. First a linear population model is given, based on the structure of the experimental situation, whose components are (unknown) parameters of the population of (conceptual) "true" responses. Then the conditions of the design are imposed to obtain a linear statistical model whose components involve the parameters of the population model and some defined random variables reflecting the experimental procedure and design. The derived statistical model is then used to obtain expected mean squares in the analyses of variance. A second illustration of the general methodology is given for a more complex example originated by Vaurio and Daniel and discussed statistically by Scheffé. Some differences from Scheffé's results are noted.

INTRODUCTION

During the past ten or so years the analysis of variance of randomized experiments has been closely identified with linear models. In fact, in many recent publications, the technique of the analysis of variance is presented in such a way that its entire meaning and validity would seem to depend upon the accuracy of certain major assumptions involving a linear model all of whose random components are normally and independently distributed.

It is true that the analysis of variance is very elegantly (and suggestively) justified and interpreted in terms of such assumptions. The

* Presented at American Statistical Association meetings, Montreal, September 1954.

The authors are indebted to H. Fairfield Smith for pointing out an error in Table 5 contained in an early draft.

† Work sponsored by Wright Air Development Center, Contract Number AF33(616)-3145.

‡ Journal paper number J-2022 of the Iowa Agricultural Experiment Station, Ames, Iowa, Project 890.

emphasis on these *ad hoc* assumptions, said to underlie the analysis of variance, has tended, however, to obscure the following important items:

(i) The assumed linear model is not a causal or mechanistic or functional relationship, in the sense, for example, that the ideal gas law, $pv = RT$, is mechanistic and functional.
(ii) The normality assumptions are always false in an actual situation.
(iii) The independence assumptions often bear no relationship to the physical situation.
(iv) Many writers lean on randomization as a justification for the assumptions, but very few have examined in detail just what randomization accomplishes and how it does so.
(v) Because the models employed are often not explicitly related to the experimental situation, there has been some difficulty in deciding just what the analysis of variance measures
(vi) Because objective methods for deriving models and the properties of their components are, in general, not used, there has been considerable controversy, even for simple situations, regarding the expectations of mean squares and the selection of appropriate error terms.

We feel that any mathematical assumptions employed in the analysis of natural phenomena must have an explicit, recognizable, relationship to the physical situation. In particular, if the analysis of variance is to be generally useful in the interpretation of experimental data it is necessary that its meaning and justification should transcend the set of arbitrary assumptions which are usually put forth

Largely because explicit and objective methods for obtaining the appropriate model for a given experiment are not generally used, one finds that certain rules and results, concerning expectations[1] of mean squares and the choice of error terms, given by Mood [10; p. 349], Hald [6, p. 480], Mentzer [9], and Scheffé [13] are contradicted by Kempthorne [7; p. 574], Anderson and Bancroft [1; p. 341], Tukey [14], Cornfield [4] and Villars [17].

It is the main purpose of the present paper to illustrate a more objective procedure for the derivation (definition) of linear models to be used as a basis for the analysis of variance The emphasis is directly on the experimental situation and design. The process of randomization, implicit or preferably explicit, plays a central role.

[1] Here "expectation" is used in the standard sense of average value over an indefinitely large number of repetitions.

RELATION TO OTHER WORK

We shall attempt to outline very briefly the relation of this paper to other relevant work. The pattern of the investigations from which the present paper stemmed is a natural generalization of the methods of "finite model analysis" given by Kempthorne [7]. The central features of both the philosophy and detail are the concept of "experimental unit" and the use of randomization in experimental designs. The dependence on the classic and extensive work of Fisher and Yates described in texts on experimental design and on the work of Neyman *et al.* [11] will be evident.

The results given in Table 2 (which involve a so-called finite population correction factor) valid for a two-factor experiment in a completely randomized design under the restrictive conditions that unit-treatment interactions are negligible correspond formally to formulas given by Tukey [14] in 1949, Cornfield [4] in 1953, and Bennett and Franklin [2] in 1954. It should be pointed out, however, that the general results, without the restrictive assumption of negligible unit-treatment interactions, given in Table 4, do not correspond. This would indicate a difference in approach in the present work. In order that confusion may not arise as to the logical basis of the various formulas which look alike the following remarks are presented.

As we understand it, Tukey's [14] results are based on an assumed linear model Tukey made explicit the idea of independent sampling of rows and columns, recognized that the interaction term is not independent of main effect terms and deduced expected mean squares which include results for random, mixed, and fixed factors. Cornfield's [4] results are based on a well-defined sampling approach, and are applicable to the situation where a population is divisible into an $R\times C$ array of primary sampling units with a constant number of sampling units within each primary unit. A sample is obtained from the intersections of r rows and c columns taken at random with n sampling units for each selected primary Bennett and Franklin's [2] results appear to have a basis similar to Tukey's.

In view of the facts that:

(1) there is no explicit attention in these particular citations to the "experimental design," by which we mean (following Fisher, Yates, Cochran, etc.) the permissible randomization patterns for the application of treatments to experimental units,[3]

[3] The experimental unit (following standard usage) is not restricted to be a physical entity such as a plot of land. It may be more abstract such as a unit of time or a set of conditions. Operationally it is that entity to which treatments are assigned at random.

(2) these authors do not give a linear model whose properties are consequences of the choice of experimental material and of treatments and of the allocation of treatments to experimental material at random;

(3) the general results exemplified in Table 4 are different;

it seems clear that the approach given in the present paper is different from that of Tukey [14], Cornfield [4], or Bennett and Franklin [2].

The work reported in this paper is part of an investigation (mainly supported by Wright Air Development Center) into the meaning and derivation of linear models and their use in the analysis of randomized experiments. Other results have been given in [21].

THE DEVELOPMENT OF A LINEAR MODEL

To illustrate our view on the meaning and derivation of a linear statistical model we will consider in this section a simple experimental situation and design, stated in general terms, under certain idealized conditions

Suppose we have two factors, $\mathcal{A}$ and $\mathcal{B}$ having A and B levels respectively, which we wish to study with respect to a given population of P experimental units Suppose that the experiment consists of selecting at random

a levels of $\mathcal{A}$,

b levels of $\mathcal{B}$,

$p = rab$ experimental units,

and applying the selected ab treatment combinations at random to the selected units so that each selected combination appears on r units This describes a two factor experiment in a completely randomized design and should not be confused with a randomized block type design.

The situation as described is rather general in that if $A=a$, $B=b$ then we have a "fixed model" situation, if $A\gg a$, $B\gg b$ then we have a "random model" situation; if $A=a$, $B\gg b$ then we have a "mixed model" situation, etc.[3]

Before going on to a derivation of the model for this situation we digress to discuss briefly the notion of experimental error. One might describe it, roughly, as the failure of repeated observations, made

[3] We use the symbol "$\gg$" to denote "much larger than."

under similar conditions, to be identical. Causes for this might be classified as

(i) unit errors—due to failure of different experimental units to yield identically under the same conditions of treatment;
(ii) treatment errors—due to inability to duplicate a treatment and conditions of its application exactly;
(iii) measurement errors—due to failure of repeated measurements on the same physical situation to correspond exactly.

This categorisation is neither rigorous nor exhaustive, but is suggestive of a real distinction between types of errors.

The unit errors must be regarded as fixed quantities associated with particular experimental units. On the other hand, the treatment and measurement (henceforth called technical) errors may often be represented mathematically as random variables.

To simplify our discussion in this section we will assume that all technical errors are negligible. We can then at once conceive of the existence of a number which would be the response if a particular treatment combination were applied to a specified experimental unit Thus if

$$i = 1, 2, \cdots, A \text{ denotes the level of } \mathcal{A};$$

$$j = 1, 2, \cdots, B \text{ denotes the level of } \mathcal{B};$$

$$k = 1, 2, \cdots, P \text{ denotes the experimental unit};$$

then we conceive of a number Y_{ijk} which would be the true response if the kth unit were subjected to the treatment combination consisting of the ith level of $\mathcal{A}$ and the jth level of $\mathcal{B}$. We note that the set $\{Y_{ijk}\}$ represents the conceivable totality of knowledge which might be experimentally acquired in this situation.

The following relationship is an algebraic identity:

$$Y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_k + n_{ijk},$$

where the quantities on the right have the following definition and physical interpretation: (we use the usual dot convention to denote means)

$\mu = Y_{\cdots}$ is the conceptual over-all mean response from all treatment combinations on all experimental units;

$a_i = (Y_{i\cdot\cdot} - Y_{\cdots})$ is the difference between the mean response from all units when subjected to the ith level of $\mathcal{A}$ in combination with each level of $\mathcal{B}$, and μ. We call a_i the main effect (or simply effect) of the ith level of $\mathcal{A}$;

$b_j = (Y_{\cdot j \cdot} - Y_{\cdots})$ is, similarly, the effect of the jth level of $\mathcal{B}$;

$(ab)_{ij} = (Y_{ij\cdot} - Y_{\cdot j \cdot} - Y_{i\cdot\cdot} + Y_{\cdots})$ is the difference between the effect of the ith level of $\mathcal{A}$ at the jth level of $\mathcal{B}$ and the effect of the ith level of $\mathcal{A}$ averaged over all the levels of $\mathcal{B}$. We call $(ab)_{ij}$ the interaction effect (or simply interaction) of the ith level of $\mathcal{A}$ and the jth level of $\mathcal{B}$;

$e_k = (Y_{\cdot\cdot k} - Y_{\cdots})$ is the difference between the mean response from all treatment combinations on the kth experimental unit, and μ. Thus e_k measures the deviation of the kth unit from the average of all the units, with respect to the average of treatment combinations. We call e_k the additive error of the kth unit;

$n_{ijk} = (Y_{ijk} - Y_{ij\cdot} - Y_{\cdot\cdot k} + Y_{\cdots})$ measures the difference between the deviation of the kth experimental unit from the average with respect to treatment combination (ij), and e_k. We call n_{ijk} the interactive error of the kth unit and treatment (ij).

We refer to the above relationship as the population model. It is based on definition and does not imply a mechanistic or causal picture —in fact, the various "levels" of the factors are not distinguished quantitatively in the statement of the model. We call particular attention to the fact that the definitions of the main effects of each factor depend in general on the levels of the other factor, and on the experimental units, which are included in the experimental situation and design. It is only when our scale of measurement is such that the factors and experimental units are mutually additive (i.e., all interactions are zero) that the preceding sentence is too restrictive.

To simplify the initial presentation somewhat we will now make the assumption that the n_{ijk} are all zero. (The assumption is not trivial, especially since there is no structuring of the experimental units, but even if untrue, it will not, in many cases, affect the interpretation of the analysis of variance too heavily, as will be evident from the following section.) Our population model then becomes

$$Y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_k.$$

We note that, by definition,

$$\sum_i a_i = \sum_j b_j = \sum_i (ab)_{ij} = \sum_j (ab_{ij}) = \sum_k e_k = 0.$$

Now, in the actual experiment we observe a randomly selected subset of the $\{Y_{ijk}\}$. Let the indexes $i^* = 1, 2, \cdots, a$ and $j^* = 1, 2, \cdots, b$ denote the randomly selected levels of $\mathcal{A}$ and $\mathcal{B}$ respectively in order of their selection. Thus, for example, $i^* = 1$ corresponds to some value of $i = i_1$ We make the convention, however, that if $A = a$ then i^* and i are the same index and similarly for the case $B = b$.

Let $x_{i^*j^*f}$ denote the observation of the fth replicate to which treatment combination (i^*j^*) is applied, where $f=1, 2, \cdots, r$ for each value of (i^*j^*) We note that to each (i^*j^*f) there corresponds a particular experimental unit, i e., some value of the index k. To write an explicit model for the $x_{i^*j^*f}$, it is convenient to introduce some additional definitions and notation.

Let

$$
\begin{aligned}
\alpha_i^{i^*} &= \text{if } i^* \text{ corresponds to } i,\\
&= 0 \text{ otherwise;}\\
\beta_j^{j^*} &= 1 \text{ if } j^* \text{ corresponds to } j,\\
&= 0 \text{ otherwise;}\\
\rho_k^{i^*j^*f} &= 1 \text{ if } (i^*j^*f) \text{ corresponds to } k,\\
&= 0 \text{ otherwise.}
\end{aligned}
$$

The α's and β's are similar to the dummy variables used by Cornfield [3] in sampling theory. The ρ's are random variables which specify how selected treatment combinations are assigned to experimental units. The quantities can be treated as random variables because random methods of selection and allocation are employed. From the design of the experiment, certain distributional properties of these quantities are easily obtained, for example.

$$\text{Prob } (\alpha_i^{i^*} = 1) = \frac{1}{A};$$

$$\text{Prob } (\alpha_i^{i^*} = 1, \alpha_{i'}^{i^{*\prime}} = 1) = \frac{1}{A(A-1)}, \qquad i \neq i', i^* \neq i^{*\prime},$$

the $\{\alpha_i^{i^*}\}$, $\{\beta_j^{j^*}\}$. $\{\rho_k^{i^*j^*f}\}$ are groupwise independent, etc.

An explicit model for $x_{i^*j^*f}$, under the simplifying assumption that there is unit treatment additivity, i.e , all $n_{ijk}=0$, is then given by

$$x_{i^*j^*f} = \mu + \sum_i \alpha_i^{i^*} a_i + \sum_j \beta_j^{j^*} b_j + \sum_{ij} \alpha_i^{i^*}\beta_j^{j^*}(ab)_{ij} + \sum_k \rho_k^{i^*j^*f} e_k.$$

This relationship we shall refer to as the statistical model. It is of interest to note that the population model derives from the experimental situation, and the statistical model is then obtained by imposing the conditions of the design of the experiment. The random variables in the statistical model are, of course, the $\alpha_i^{i^*}$, $\beta_j^{j^*}$ and $\rho_k^{i^*j^*f}$, which take on the values 0 and 1 with known probabilities. All other quantities in the model are fixed, unknown, parameters.

This development of a model is rather general. For example, if $A=a$, $B\gg b$ then we have a "mixed model" situation with factor $\mathcal{A}$ fixed and factor $\mathcal{B}$ random. Then we take i and i^* to be identical indexes Thus

$$\sum_i \alpha_i^{i^*} a_i = a_{i^*}, \quad \text{and} \quad \sum_i \alpha_i^{i^*}(ab)_{ij} = (ab)_{i^*j}.$$

Hence, the mixed model becomes

$$x_{i^*j^*f} = x_{ij^*f} = \mu + a_i + \sum_j \beta_j^{j^*} b_j + \sum_j \beta_j^{j^*}(ab)_{ij} + \sum_k \rho_k^{ij^*f} e_k.$$

THE ANALYSIS OF VARIANCE

The algebraic structure of the analysis of variance for this design is given in Table 1 We use the usual dot convention to denote means.

TABLE 1

ANALYSIS OF VARIANCE

$\mathcal{A}$	$(a-1)$	$A'=br\sum_{i^*}(x_{i^*\cdot\cdot}-x_{\cdot\cdot\cdot})^2$	$A^*=A'/(a-1)$
$\mathcal{B}$	$(b-1)$	$B'=ar\sum_{j^*}(x_{\cdot j^*\cdot}-x_{\cdot\cdot\cdot})^2$	$B^*=B'/(b-1)$
$\mathcal{A}\times\mathcal{B}$	$(a-1)(b-1)$	$I'=r\sum_{i^*j^*}(x_{i^*j^*\cdot}-x_{i^*\cdot\cdot}$ $-x_{\cdot j^*\cdot}+x_{\cdot\cdot\cdot})^2$	$I^*=I'/(a-1)(b-1)$
Residual	$ab(r-1)$	$R'=\sum_{i^*j^*f}(x_{i^*j^*f}-x_{i^*j^*\cdot})^2$	$R^*=R'/ab(r-1)$
Total	$abr-1$	$G'=\sum_{i^*j^*f}(x_{i^*j^*f}-x_{\cdot\cdot\cdot})^2$	

Using the derived statistical model and some distributional properties of the $\{\alpha_i^{i^*}\}$, $\{\beta_j^{j^*}\}$ and $\{\rho_k^{i^*j^*f}\}$, the expectations of the mean squares can be readily obtained The algebra is exemplified in the Appendix. The results are given in Table 2 The following notational definitions are used.

$$\sigma_a^2 = \frac{1}{(A-1)}\sum_i a_i^2$$

$$\sigma_b^2 = \frac{1}{(B-1)}\sum_j b_j^2$$

$$\sigma_{ab}^2 = \frac{1}{(A-1)(B-1)}\sum_{ij}(ab)_{ij}^2$$

$$\sigma_e^2 = \frac{1}{(P-1)}\sum_k e_k^2$$

TABLE 2

EXPECTATIONS OF MEAN SQUARES UNDER ASSUMPTION OF UNIT-TREATMENT ADDITIVITY

Mean Square	Expectation of Mean Square
A^*	$\sigma_e^2+\frac{(B-b)}{B}r\sigma_{ab}^2+rb\sigma_a^2$
B^*	$\sigma_e^2+\frac{(A-a)}{A}r\sigma_{ab}^2+ra\sigma_b^2$
I^*	$\sigma_e^2+r\sigma_{ab}^2$
R^*	σ_e^2

We see from this analysis that randomization "controls" the unit errors, in a statistical sense, and enables the estimation of certain defined "components of variance," as well as the error of estimates of contrasts among treatments actually employed. Furthermore, tests of significance are made possible by use of randomization tests.

We employed the assumption of additivity of treatments and experimental units. If the unit-treatment interactions are not negligible then as will be seen in the following section the general effect will be to underestimate significance and overestimate errors.

We now turn our attention to the specification of "proper error terms." The present use of the phrase is exemplified by the following

> E_a^* is said to be the proper error term for $\mathcal{A}$ main effects if E_a^* is a linear combination of the analysis of variance mean squares such that $E(E_a^*)=E(A^*)-rb\sigma_a^2$, where by $E(E_a^*)$ we mean the expectation of E_a^*.

The random variable E_a^* will then have the properties that

(i) an unbiased estimate of σ_a^2 is given by a multiple of $(A^*-E_a^*)$

(ii) a meaningful "F-type" statistic, for the test of significance of $\mathcal{A}$ main effects is given by A^*/E_a^*;

(iii) a multiple of E_a^* is an unbiased estimate of the variance of the estimate of any linear contrast of the effects of levels of $\mathcal{A}$ actually used.

The appropriate multiplier in (i) and (iii) will depend on A, B, a, b, and the particular contrast.

The proper error terms, with unit-treatment additivity, for the general situation under consideration are given in Table 3.

TABLE 3

PROPER ERROR TERMS WITH UNIT-TREATMENT ADDITIVITY

Category	Proper Error Term
$\mathcal{A}$ effects	$I^* - \frac{b}{B}(I^* - R^*) = \frac{(B-b)}{B} I^* + \frac{b}{B} R^*$
$\mathcal{B}$ effects	$I^* - \frac{a}{A}(I^* - R^*) = \frac{(A-a)}{A} I^* + \frac{b}{A} R^*$
$\mathcal{A} \times \mathcal{B}$ interactions	R^*

It is easy now to examine some common, important, special cases.

(a) Fixed Model: $A = a$, $B = b$.

The proper error term for each of $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{A} \times \mathcal{B}$ is R^*.

(b) Random Model: $A \gg a$, $B \gg b$.

Under these conditions a/A and b/B approach zero and the proper error term for $\mathcal{A}$ and also for $\mathcal{B}$ is I^*.

(c) Mixed Model: $A \gg a$, $B = b$.

Here $\mathcal{A}$ is random and $\mathcal{B}$ is fixed. The proper error term for the random factor $\mathcal{A}$ is R^* The proper error term for the fixed factor $\mathcal{B}$ is I^*.

Cases (a) and (b) correspond in a certain sense to Models I and II, respectively, of Eisenhart [5]

RESULTS UNDER GENERAL CONDITIONS

The analysis of the two-factor completely randomized experiment given in the preceding sections was based on the simplifying assumption that the interactions of treatments with experimental units were negligible, i.e., that $n_{ijk} = 0$ for all i, j, k. This assumption was employed to simplify the exposition of the general method, which is, of course, independent of the assumption above. In this section we indicate the more general results.

We note that $\sum_{ij} n_{ijk} = \sum_k n_{ijk} = 0$.

The general statistical model is

$$x_{i^*j^*f} = \mu + \sum_i \alpha_i^{i^*} a_i + \sum_j \beta_j^{j^*} b_j + \sum_{ij} \alpha_i^{i^*} \beta_j^{j^*} (ab)_{ij} + \sum_k \rho_k^{i^*j^*f} e_k$$
$$+ \sum_{ijk} \alpha_i^{i^*} \beta_j^{j^*} \rho_k^{i^*j^*f} n_{ijk}.$$

The analysis of variance is as given in Table 1. The expectations of mean squares under the more general conditions are given in Table 4.

In addition to the notational definitions given previously we use

$$\sigma_n^2 = \frac{1}{AB(P-1)} \sum_{ijk} n_{ijk}^2; \qquad Q_{an}^2 = \frac{1}{(A-1)(P-1)} \sum_{ik} n_{i\cdot k}^2;$$

$$Q_{bn}^2 = \frac{1}{(B-1)(P-1)} \sum_{jk} n_{\cdot jk}^2; \quad Q_{abn}^2 = \frac{1}{(A-1)(B-1)(P-1)} \cdot \sum_{ijk} (n_{ijk} - n_{i\cdot k} - n_{\cdot jk})^2.$$

TABLE 4

EXPECTATIONS OF MEAN SQUARES
(MORE GENERAL CONDITIONS)

Mean Square	Expectation of Mean Square
A^*	$\sigma_n^2 + \sigma_e^2 = \frac{(B-b)}{B} r\left[\sigma_{ab}^2 - \frac{1}{P} Q_{abn}^2\right] + rb\left[\sigma_a^2 - \frac{1}{P} Q_{an}^2\right]$
B^*	$\sigma_n^2 + \sigma_e^2 + \frac{(A-a)}{B} r\left[\sigma_{ab}^2 - \frac{1}{P} Q_{abn}^2\right] + ra\left[\sigma_b^2 - \frac{1}{P} Q_{bn}^2\right]$
I^*	$\sigma_n^2 + \sigma_e^2 + r\left[\sigma_{ab}^2 - \frac{1}{P} Q_{abn}^2\right]$
R^*	$\sigma_n^2 + \sigma_e^2$

It will be seen from Table 4 that, in general, unbiased estimates of σ_a^2, σ_b^2, and σ_{ab}^2 cannot be obtained from the analysis of variance mean squares, and that proper error terms do not exist. However, the "bias" is of order $1/P$, where P is the number of experimental units in the population.

Thus, for example, if E_a^* is the error term for $\mathcal{A}$ effects given in Table 3, then the bias in using

$$\frac{(A^* - E_a^*)}{rb}$$

as an estimate of σ_a^2 will be $[(-(1/P)Q_{an}^2]$. In many cases this bias will be negligible.

A MORE COMPLEX EXAMPLE

To apply these notions in a more complex situation and to contrast the procedures with those given by other writers, we will consider an example given by Vaurio and Daniel [16], and discussed in detail by Scheffé [13].

The example concerns the investigation of several sources of vari-

ation in the pack life of cans in a prune pack test. The fundamental units as manufactured and treated are long coils of thin steel. Important variation is expected from coil to coil and also from end to end within each coil, the latter variation being due, apparently, to systematic variation in the preparation of each coil. The experiment is intended to investigate these sources of variation as well as that due to several different levels of annealing (a treatment applied to an entire coil) on the pack life of cans, under certain standard conditions

Suppose then that there are A levels of anneal to be investigated, and that L *definite* locations (say head, middle, tail, etc.) are distinguished on all coils

The experiment involves the (random) selection of Ac coils from some large (perhaps conceptual) population of C such coils. We then apply each of the A anneal levels to c of the selected coils, using random allocation

From each location on every treated coil are selected at random s specimens of metal These are made up into cans, filled, and time to failure observed

We will suppose for simplicity that all locations on each coil are of equal size, say each containing S possible specimens

Let

$i=1, 2, \cdots, A$ denote the anneal level,
$j=1, 2, \cdots, C$ denote the coil in the population of coils;
$k=1, 2, \cdots, L$ denote location;
$m=1, 2, \cdots, S$ denote the specimen for a given location on a given coil;
$j^*=1, 2, \cdots, c$ denote the coils as randomly allocated to each anneal level (i e , for each value of i, j^* takes on values from 1 to c).
$m^*=1, 2, \cdots, s$ denote the randomly selected specimens for each location on each selected coil, i e , for each value of (ij^*k), m^* takes on values from 1 to s.

Now let us imagine that the conditions of test, of measurement, and of application of treatment are precisely standardized Then we can conceive of a number Y_{ijkm} which would be the pack life of the mth specimen at the kth location on the jth coil when the coil has been subjected to the ith anneal level Then we can write the identity

$$Y_{ijkm} = \mu + \alpha_i{}^A + \alpha_j{}^C + \alpha_{ij}{}^{AC} + \alpha_k{}^L + \alpha_{ik}{}^{AL} + \alpha_{jk}{}^{CL} + \alpha_{ijk}{}^{ACL} + a_{jkm}{}^S + \alpha_{ijm}{}^{AS},$$

when

$\mu = Y.\,..$ is the "true" over-all average pack life;

$\alpha_i^A = (Y_i\ldots - Y\ldots.)$ is the effect of the ith anneal level;

$\alpha_j^C = (Y\,_j.. - Y.\quad)$ is the effect of the jth coil;

$\alpha_{ij}^{AC} = (Y_{ij}.. - Y_i\ldots - Y\,_j..+Y\;.\;.)$ is the interaction of the ith anneal and jth coil;

$\alpha_k^L = (Y.._k - Y.\quad.)$ is the effect of the kth location;

$\alpha_{ik}^{AL} = (Y_i._k. - Y_i.\quad - Y.._k.+Y\;..\;)$ is the interaction of the ith anneal and the kth location;

$\alpha_{jk}^{AL} = (Y._{jk} - Y._j\;. - Y.\;_k + Y\ldots.)$ is the interaction of the jth coil and the kth location;

$\alpha_{ijk}^{ACL} = (Y_{ijk} - Y_{ij} - Y_i._k.+Y_i\;.. - Y._{jk} + Y._j..+Y.._k. - Y\ldots\;)$ is the interaction of the ith anneal, jth coil and kth location;

$\alpha_{jkm}^S = (Y._{kjm} - Y._{jk}\;)$ is the within coil and location "error" associated with mth specimen;

$\alpha_{ijkm}^{AS} = (Y_{ijkm} - Y_{ijk} - Y\,_{jkm} + Y._{jk}.)$ is the interactive "error" of the ith anneal with the mth specimen within the kth location on the jth coil.

To simplify the algebraic treatment somewhat we will assume that the interaction terms $\{\alpha_{ijk}^{ACL}\}$ and $\{\alpha_{ijkm}^{AS}\}$ are negligible. Scheffé assumes these interactions as well as the $\{\alpha_{ij}^{AC}\}$ are negligible We shall however carry the anneal-coil interactions to observe their effect on the analysis. Thus our population model becomes

$$Y_{ijkm} = \mu + \alpha_i^A + \alpha_j^C + \alpha_{ij}^{AC} + \alpha_{ik}^{AL} + \alpha_{jk}^{CL} + \alpha_{jkm}^S.$$

Of course we know that the conditions of test for cans are not entirely standard and hence, if we tested the can from the mth specimen in the kth location on the jth coil which was treated with the ith anneal level we would not observe Y_{ijkm}. We will assume here that we may treat the conceptual observable, y_{ijkm}, as a random variable such that

$$y_{ijkm} = Y_{ijkm} + \epsilon_{ijkm},$$

where the ϵ_{ijkm} are mutually uncorrelated random variables, with mean zero and constant variance σ^2.

Let $x_{ij^*km^*}$ represent the observation obtained from the m^*th specimen at the kth location from the j^*th coil which was treated with the ith anneal level. Define the random variables

$$\rho_j^{ij^*} = 1 \text{ if } (ij^*) \text{ corresponds to } j$$
$$= 0 \text{ otherwise}$$
$$\xi_{ij^*km}^{ij^*km^*} = 1 \text{ if } (ij^*km^*) \text{ corresponds to } (ij^*km)$$
$$= 0 \text{ otherwise}$$

Then

$$\text{Prob } (\rho_j^{ij^*} =)1 = \frac{1}{C};$$

$$\text{Prob } (\rho_j^{ij^*} = 1, \rho_j^{i'j^{*\prime}} = 1) = \frac{1}{C(C-1)}; \qquad (ij^* \neq i'j^{*\prime})$$

$$\text{Prob } (\xi_{ij^*km}^{ij^*km^*} = 1) = \frac{1}{S};$$

the $\{\rho_j^{ij^*}\}$ and $\{\xi_{ij^*km}^{ij^*km^*}\}$ are independent, etc.

We can then write the following model for the experimental observable:

$$x_{ij^*km^*} = \mu + \alpha_i^A + \sum_j \beta_j^{ij^*}\alpha_j^C + \sum_j \rho_j^{ij^*}\alpha_{ij}^{AC} + \alpha_k^L + \alpha_{ik}^{AL}$$
$$+ \sum_j \rho_j^{ij^*}\alpha_{jk}^{CL} + \sum_{jm} \rho_j^{ij^*}\xi_{ij^*km}^{ij^*km^*}(\alpha_{jkm}^S + \epsilon_{ijkm})$$

If we put

$$a_{ij^*}^C = \sum_j \rho_j^{ij^*}\alpha_j^C;$$
$$a_{ij^*}^{AC} = \sum_j \rho_j^{ij^*}\alpha_{ij}^{AC};$$
$$a_{ij^*k}^{CL} = \sum_j \rho_j^{ij^*}\alpha_{jk}^{CL};$$
$$a_{ij^*km^*}^S = \sum_{jm} \rho_j^{ij^*}\xi_{ij^*km}^{ij^*km^*}\alpha_{jkm}^S;$$
$$\epsilon^*_{ij^*km^*} = \sum_{jm} \rho_j^{ij^*}\xi_{ij^*km}^{ij^*km^*}\epsilon_{ijkm};$$

then,

$$x_{ij^*km^*} = \mu + \alpha_i^A + a_{ij^*}^C + a_{ij^*}^{AC} + \alpha_k^L + \alpha_{ik}^{AL} + \alpha_{ij^*k}^{CL}$$
$$+ a_{ij^*km^*}^S + \epsilon^*_{ij^*km^*}.$$

This last relationship is notationally similar to the linear model assumed by Scheffé [13], but some important differences in properties exist, for example Scheffé assumes that the random variables $a_{ij^*k}^{CL}$ are mutually independent, while from our derivation of the model it follows that

$$\sum_k a_{ij^*k}^{CL} = \sum_k \left(\sum_j \rho_j^{ij^*} \alpha_{jk}^{CL} \right) = \sum_j \rho_j^{ij^*} \left(\sum_k \alpha_{jk}^{CL} \right) = 0.$$

We give in Table 5 the expectations of the analysis of variance mean squares obtainable from the statistical model we have derived. The following notational definitions are used:

$$\sigma_A^2 = \frac{1}{(A-1)} \sum_i (\alpha_i^A)^2; \quad \sigma_C^2 = \frac{1}{C-1} \sum_j (\alpha_j^C)^2;$$

$$\sigma_L^2 = \frac{1}{L-1} \sum_k (\alpha_k^L)^2; \quad \sigma_{AC}^2 = \frac{1}{A(C-1)} \sum_{ij} (\alpha_{ij}^{AC})^2,$$

$$\sigma_{AL}^2 = \frac{1}{(A-1)(L-1)} \sum_{ik} (\alpha_{ik}^{AL})^2;$$

$$\sigma_{CL}^2 = \frac{1}{(C-1)(L-1)} \sum_{jk} (\alpha_{jk}^{CL})^2;$$

$$\sigma_S^2 = \frac{1}{CL(S-1)} \sum_{jkm} (\alpha_{jkm}^S)^2; \quad \sigma^2 = E(\epsilon_{ijkm}^2).$$

For each of the above definitions, with the exception of σ_{AC}^2, the divisor is selected as follows. if the component is based on the sum of squares of N quantities which are subject to n linearly independent linear relations, then the divisor is $(N-n)$. According to this criterion we could define

$$\sigma_{AC}^{2\prime} = \frac{1}{(A-1)(C-1)} \sum_{ij} (\alpha_{ij}^{AC})^2$$

for the anneals by coils interaction component; we have defined σ_{AC}^2 with a divisor of $A(C-1)$ because this fits a pattern which simplifies the statement of more general results. If the reader prefers, he may replace σ_{AC}^2 in Table 5 by

$$\frac{(A-1)}{A} \sigma_{AC}^{2\prime}.$$

TABLE 5

EXPECTATIONS OF MEAN SQUARES

Due to	d.f.	M. S.	Expected Mean Square
Anneals ($\mathcal{A}$)	$(A-1)$	A^*	$\sigma^2+\frac{(S-s)}{S}\sigma_S^2\left(1-\frac{Ac}{(A-1)C}\right)Ls\sigma_{AC}^2+Ls\sigma_C^2+cLs\sigma_A^2$
Coils ($\mathcal{C}$)	$A(c-1)$	C^*	$\sigma^2+\frac{(S-s)}{S}+Ls\sigma_{AC}^2+Ls\sigma_C^2$
Locations ($\mathcal{L}$)	$(L-1)$	L^*	$\sigma^2+\frac{(S-s)}{S}\sigma_S^2+\frac{(C-Ac)}{C}s\sigma_{CL}^2+Acs\sigma_L^2$
$\mathcal{A}\times\mathcal{L}$	$(A-1)(L-1)$	I_{AL}^*	$\sigma^2+\frac{(S-s)}{S}\sigma_S^2+s\sigma_{CL}^2+cs\sigma_{AL}^2$
$\mathcal{C}\times\mathcal{L}$	$A(c-1)(L-1)$	I_{CL}^*	$\sigma^2+\frac{(S-s)}{S}\sigma_S^2+s\sigma_{CL}^2$
Specimens ($\mathcal{S}$)	$AcL(s-1)$	S^*	$\sigma^2+\sigma_S^2$

The coefficient of $\sigma_{AC}^{2\prime}$ in the expected mean square for anneals is

$$\left(1-\frac{1}{A}-\frac{c}{C}\right)Ls$$

and for coils is $((A-1)/A)Ls$

In this example $S \gg s$, and we treat the coils actually used as a small random selection from a large population so that $C \gg Ac$. Hence,

$$\frac{(S-s)}{S},\quad \left(1-\frac{Ac}{(A-1)C}\right)\quad \text{and}\quad \left(\frac{C-Ac}{C}\right)$$

may be taken as equal to 1 Thus we see that independent of any assumptions regarding the anneal-coil interactions, we have that $E(C^*)=E(A^*)-cLs\sigma_A^2$, and so C^* is an appropriate error term for the evaluation of the effects of the anneal levels in this experiment independent of any assumptions about the $\{\alpha_{ij}^{AC}\}$.

From our analysis we would thus list the following "proper error terms":

Component	*Proper Error Term*
σ_A^2	C^*
$\sigma_C^2+\sigma_{AC}^2$	S^*
σ_L^2	I_{CL}^*
σ_{AL}^2	I_{CL}^*
σ_{CL}^2	S^*

For the same example, Scheffé uses[4] the assumed linear model

$$x_{ij^*km} = \mu + \alpha_i^{A} + a_{ij^*}^{C} + \alpha_k^{L} + \alpha_{ik}^{AL} + \alpha_{ij^*k}^{CL} + a_{ij^*km}^{S},$$

where the α's are taken to be fixed constants, and the a's are assumed to be normally and independently distributed random variables, with means zero and constant variances within groups. We reproduce in Table 6 Scheffé's expectations of mean squares.

TABLE 6

EXPECTATIONS OF MEAN SQUARES, FROM SCHEFFÉ [13]

Mean Square	Expectation
A^*	$\sigma_S^2 + s\sigma_{CL}^2 + Ls\sigma_C^2 + cLs\sigma_A^2$
C^*	$\sigma_S^2 + s\sigma_{CL}^2 + Ls\sigma_C^2$
L^*	$\sigma_S^2 + s\sigma_{CL}^2 + Acs\sigma_L^2$
I_{AL}^*	$\sigma_S^2 + s\sigma_{CL}^2 + cs\sigma_{AL}^2$
I_{CL}^*	$\sigma_C^2 + s\sigma_{CL}^2$
S^*	σ_S^2

In view of the differences between Tables 5 and 6 the reader may ask "who is right?" There can be no simple answer to this question. All one can do is to use methodology with the most acceptable assumptions and the choice is between the logic of procedures rather than simply between results. Our general views on the basis for choice of a statistical technique are given above and below; our aim in this section has simply been to present the details of our approach for an example to which another point of view had been applied

CONCLUDING REMARKS

The subjects of this concluding section fall into three categories: first, advantages as we see them of the procedure which has been illustrated heretofore in this paper; second, some views on problems of distribution; third, some brief remarks on where and why the various types of models are appropriate.

The following advantages accrue to the procedure suggested in this paper:

(i) An explicit connection is given between the physical situation and the components of the model. Thus there is no doubt that the statistical model is definitional rather than mechanistic (or in Tukey's [15] language, is tangential rather than functional).

[4] We have adjusted Scheffé's notation somewhat to allow easy comparison with the present results.

(ii) It is made clear that the meaning of the main effect of a certain level of a factor depends not only on the other levels of that factor but also, in general, on the levels of all other factors which are involved.

(iii) An explicit definition is given of interaction, and hence some insight is provided into the fact that the much used assumptions of additivity[5] of factors and of additivity of treatments with experimental material depend not only on the mechanics of the experiment but also on the scale of measurement.

(iv) This approach forces on the experimenter and the statistician the fact that it is possible to draw inferences statistically only about populations from which samples are drawn according to probability considerations—i.e., it is the design and procedure of the experiment that determines the scope of the statistical inference, and extensions of such inferences to wider circumstances cannot be assessed probability-wise

(v) The procedure we have used is hinged to the experimental situation and design rather than around a set of arbitrary assumptions necessary to justify some esoteric distribution theory. Thus attention is focussed on the physical interpretations and distributional assumptions are relegated to their proper role of serving as a guide in obtaining mathematical approximations.

(vi) Finally, general formulas may be obtained by this procedure which are valid for all possible situations of which "all fixed" and "all random" models are extreme special cases. On the other hand, the usual approach treats fixed and random models quite differently and gives no clue on how to write a model for cases where the factors are neither fixed nor random.

There are many distributional problems associated with analysis of variance tests of significance which are not hinged on a priori normality assumptions. (There are also problems of distribution even when the whole battery of normal theory assumptions is made.) Two viewpoints suggest themselves. One can use the approach illustrated in this paper to derive the appropriate model and specify certain inherent properties of the components of the model, and then "superimpose" a sufficient number of assumptions in order to obtain so-called "exact" tests. Alternatively, one may rely on the fundamental notion of randomization tests, and use as approximations to randomization distributions certain procedures suggested by normal theory. (See Welch [18], Pitman [12], Kempthorne [7, 8], Wilk [19, 20].)

[5] We use the word "additivity" in the sense of "negligible interaction effects."

We prefer the latter viewpoint of mathematical approximation to the former of statistical assumptions because (i) it is less likely to obscure the tenuous relation between the physical situation and mathematical or statistical abstractions; and (ii) it faces honestly the fact that more than desire for mathematical elegance is required to validate statistical assumptions. (These remarks are not intended to depreciate normal theory procedures, which we believe serve very importantly as general suggestive guides in statistical techniques)

The final paragraphs concern the appropriateness of various types of models It is clear that the model must reflect the experiment as actually carried out, and not any desire to make a broad scientific generalization. This, however just pushes the question back to the experimental design. The following experimental situations may be distinguished.

(i) Even though a factor has a well-defined (say quantitatively) array (or continuum) of possible levels we choose to examine a "representative" set, e g., in studying the behavior of a process in a temperature range say from 100–125°C we choose to study it at 100°C, 105°C, $\cdots$, 125°C This factor, temperature, is then to be regarded as fixed. We may of course have prior information which enables us to interpolate between the temperatures actually used The analysis of variance may then be used to assess the adequacy or reliability of the experiment and perhaps also in connection with functional analyses by attempting to find scales on which interactions become negligible.

(ii) A factor may have a definite number of qualitatively distinguishable categories and we have the interest and resources to study all of these. For example, we may test a definite number of types of vacuum tubes for a given application Again, this factor is regarded as fixed. In this situation, functional analysis may be more difficult or of little interest

(iii) Resources may not permit us to examine all possible "levels" of a given factor, and we examine a randomly selected subset to get an estimate of the variability of these levels. For example, we may wish to test the lethality of an antibiotic with respect to a bacterium which has 50 genetic strains. Our experiment involves the random selection of 10 of these strains. This is a case where the general development we have given would be applied. One might refer to this factor as "semi-random" in order to preserve the usual terminology.

(iv) We may have a large population of operationally distinct entities (such as pilots, carloads of oil, blocks of land, etc.) about which we wish to make some inference, and our experiment involves the random selection of a comparatviely small sample of the large number of "levels" of this "factor." Such a factor is then regarded as random.

(v) There are some circumstances under which it may be useful to consider the levels of a random factor actually used as though they were the levels of a fixed factor (with a corresponding redefinition of main effects and interactions), but there appears to be no objective basis for the converse case.

APPENDIX

The object of this appendix is to exemplify the procedure by which the expectations of mean squares are obtained from the statistical model We shall outline this derivation for the $\mathcal{A}$ mean square and the residual mean square of Table 1, under the simplifying assumption of unit-treatment additivity.

Consider the residual sum of squares

$$R' = \sum_{i^*j^*f} (x_{i^*j^*f} - x_{i^*j^*})^2$$

Substituting the statistical model we find that

$$R' = \sum_{i^*j^*f} \left(\sum_k \rho_k^{i^*j^*f} e_k - \frac{1}{r} \sum_{fk} \rho_k^{i^*j^*f} e_k \right)^2$$

$$= \sum_{i^*j^*} \left[\sum_f \left(\sum_k \rho_k^{i^*j^*f} e_k \right)^2 - \frac{1}{r} \left(\sum_{fk} \rho_k^{i^*j^*f} e_k \right)^2 \right]$$

Now $(\rho_k^{i^*j^*f} \rho_k^{i^*j^*f'}) = 0$ with probability 1, for $f \neq f'$ and $(\rho_k^{i^*j^*f} \rho_{k'}^{i^*j^*f} = 0$ with probability 1, for $k \neq k'$, and any "ρ" variable is equal to zero or unity. Hence,

$$R' = \sum_{i^*j^*} \left[\sum_{fk} \rho_k^{i^*j^*f} e_k^2 - \frac{1}{r} \sum_{fk} \rho_k^{i^*j^*f} e_k^2 - \frac{1}{r} \sum_{\substack{f \neq f' \\ k \neq k'}} \rho_k^{i^*j^*f} \rho_{k'}^{i^*j^*f} e_k e_{k'} \right]$$

Taking expectations,

$$E(R') = \sum_{i^*j^*} \left[\sum_{fk} \frac{1}{P} e_k^2 - \frac{1}{r} \sum_{fk} \frac{1}{P} e_k^2 - \frac{1}{r} \sum_{\substack{f \neq f' \\ k \neq k'}} \frac{1}{P(P-1)} e_k e_{k'} \right]$$

$$= ab \left[\frac{(r-1)}{P} \sum_k e_k^2 - \frac{(r-1)}{P(P-1)} \sum_{k \neq k'} e_k e_{k'} \right].$$

But $\sum_k e_k = 0$ Hence

$$\sum_{k \neq k'} e_j e_{k'} = - \sum_k e_k^2$$

So,

$$E(R') = \frac{ab(r-1)}{P(P-1)} (P - 1 + 1) \sum_k e_k^2.$$

$$= ab(r-1)\sigma_e^2$$

Dividing by the degrees of freedom, we obtain $E(R^*) = \sigma_e^2$, as given in Table 2.

Turning now to the $\mathcal{A}$ sum of squares, in Table 1,

$$
\begin{aligned}
A' &= br \sum_{i^*} (x_{i^*\cdot} - x_{\cdots})^2 \\
&= br \sum_{i^*} \Bigg[\Big(\sum_i \alpha_i^{i^*} a_i - \frac{1}{a} \sum_{i^*i} \alpha_i^{i^*} a_i \Big) + \Big(\frac{1}{b} \sum_{j^*ij} \alpha_i^{i^*} \beta_j^{j^*} (ab)_{ij} \\
&\quad - \frac{1}{ab} \sum_{i^*j^*ij} \alpha_i^{i^*} \beta_j^{j^*} (ab)_{ij} \Big) + \Big(\frac{1}{br} \sum_{j^*fk} \rho_k^{i^*j^*f} e_k \\
&\quad - \frac{1}{abr} \sum_{i^*j^*fk} \rho_k^{i^*j^*f} e_k \Big) \Bigg]^2 \\
&= br \sum_{i^*} (A_{1i^*} + A_{2i^*} + A_{3i^*})^2, \text{ say}
\end{aligned}
$$

All products such as $A_{1i}^* A_{3i}^*$ have expected value 0. For example,

$$
\begin{aligned}
E(A_{1i^*} A_{3i^*}) = E\Big(\sum_i \alpha_i^{i^*} a_i - \frac{1}{a} \sum_{i^*i} \alpha_i^{i^*} a_i \Big) \\
\cdot \Big(\frac{1}{br} \sum_{j^*fk} \rho_k^{i^*j^*f} e_k - \frac{1}{abr} \sum_{i^*j^*fk} \rho_k^{i^*j^*f} e_k \Big),
\end{aligned}
$$

and

$$
\begin{aligned}
E\Big(\sum_i \sum_{j^*fk} \alpha_i^{i^*} a_i \rho_k^{i^*j^*f} e_k \Big) &= \sum_{ij^*fk} \frac{1}{AP} a_i e_k \\
&= 0, \text{ since } \alpha_i^{i^*} \text{ and } \rho_k^{i^*j^*f} \text{ are independent,}
\end{aligned}
$$

and $\sum_k e_k = \sum_i a_i = 0$. Similarily, the other cross products in $A_{1i}^* A_{3i}^*$ have expectation 0.

Thus,

$$
E(A') = brE\Big[\sum_{i^*} (A_{1i^*}^2 + A_{2i^*}^2 + A_{3i^*}^2) \Big].
$$

Now,

$$
\begin{aligned}
E\Big(\sum_{i^*} A_{1i^*}^2 \Big) &= E\Big[\sum_{i^*} \Big(\sum_i \alpha_i^{i^*} a_i \Big)^2 - \frac{1}{a} \Big(\sum_{i^*i} \alpha_i^{i^*} a_i \Big)^2 \Big] \\
&= E \sum_{i^*i} \alpha_i^{i^*} a_i^2 - \frac{1}{a} \sum_{i^*i} \alpha_i^{i^*} a_i^2 - \frac{1}{a} \sum_{\substack{i^* \neq i^{*\prime} \\ i \neq i'}} \alpha_i^{i^*} \alpha_{i'}^{i^{*\prime}} a_i a_{i'},
\end{aligned}
$$

by the same reasoning as in dealing with R',

$$
\begin{aligned}
&= \frac{(a-1)}{A} \sum_i a_i^2 - \frac{a(a-1)}{aA(A-1)} \sum_{i \neq i'} a_i a_{i'} \\
&= (a-1)\sigma_a^2, \text{ since } \sum_i a_i = 0.
\end{aligned}
$$

Similarly

$$E\Big(\sum_{i^*} A_{3i^*}{}^2\Big) = E\Big\{\sum_{i^*}\Big[\frac{1}{b}\sum_{j^*ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij} - \frac{1}{ab}\sum_{i^*j^*ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}\Big]^2\Big\}$$

$$= E\Big[\sum_{i^*}\Big(\frac{1}{b}\sum_{j^*ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}\Big)^2 - \frac{1}{a}\Big(\frac{1}{b}\sum_{i^*j^*ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}\Big)^2\Big]$$

We have

$$\Big[\sum_{j^*ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}\Big]^2 = \sum_{j^*}\Big[\sum_{ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}\Big]^2 + \sum_{j^*\neq j^{*\prime}}\Big[\sum_{ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}\Big]\Big[\sum_{ij}\alpha_i{}^{i^*}\beta_j{}^{j^{*\prime}}(ab)_{ij}\Big]$$

$$= \sum_{j^*}\sum_{ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}{}^2 + \sum_{j^*\neq j^{*\prime}}\sum_{\substack{i\\ j\neq j'}}\alpha_i{}^{i^*}\beta_j{}^{j^*}\alpha_i{}^{i^*}\beta_j{}^{j^{*\prime}}(ab)_{ij}(ab)_{ij'}.$$

The expectation of this is

$$\frac{b}{AB}\sum_{ij}(ab)_{ij}{}^2 - \frac{b(b-1)}{AB(B-1)}\sum_{ij}(ab)_{ij}{}^2 = \frac{b}{AB}\frac{(B-b)}{(B-1)}\sum_{ij}(ab)_{ij}{}^2.$$

Also

$$\Big[\sum_{i^*j^*}\sum_{ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}\Big]^2 = \sum_{i^*j^*}\Big[\sum_{ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}{}^2\Big] + \sum_{\substack{i^*\neq i^{*\prime}\\ j^*}}\Big[\sum_{ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}\Big]\Big[\sum_{ij}\alpha_i{}^{i^{*\prime}}\beta_j{}^{j^*}(ab)_{ij}\Big]$$

$$+ \sum_{\substack{i^*\\ j^*\neq j^{*\prime}}}\Big[\sum_{ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}\Big]\Big[\sum_{ij}\alpha_i{}^{i^*}\beta_j{}^{j^{*\prime}}(ab)_{ij}\Big] + \sum_{\substack{i^*\neq i^{*\prime}\\ j^*\neq j^{*\prime}}}\Big[\sum_{ij}\alpha_i{}^{i^*}\beta_j{}^{j^*}(ab)_{ij}\Big]\Big[\sum_{ij}\alpha_i{}^{i^{*\prime}}\beta_j{}^{j^{*\prime}}(ab)_{ij}\Big].$$

The expectation of this is

$$\Big[\sum_{ij}(ab)_{ij}{}^2\Big]\Big[\frac{ab}{AB} - \frac{a(a-1)}{A(A-1)} - \frac{b(b-1)}{AB(B-1)} + \frac{a(a-1)b(b-1)}{A(A-1)B(B-1)}\Big]$$

$$= \sum_{ij}(ab)_{ij}{}^2\frac{ab}{AB}\frac{(A-a)}{(A-1)}\frac{(B-b)}{(B-1)}$$

Combining terms we have

$$E\Big(\sum_{i^*}A_{3i^*}{}^2\Big) = \Big[\frac{a}{b^2}\frac{b(B-b)}{AB(B-1)} - \frac{1}{ab^2}\frac{ab(A-a)(B-b)}{AB(A-1)(B-1)}\Big]\Big(\sum_{ij}(ab)_{ij}{}^2\Big)$$

$$= \frac{(a-1)}{b}\frac{(B-b)}{B}\sigma_{ab}{}^2.$$

In the same way we find that

$$E(A_{3,*^2}) = \frac{(a-1)}{br}\sigma_e^2$$

Hence

$$E(A') = br\left[(a-1)\sigma_a^2 + \frac{(a-1)}{b}\frac{(B-b)}{B}\sigma_{ab}^2 + \frac{(a-1)}{br}\sigma_e^2\right]$$

$$= (a-1)\left[\sigma_e^2 + \frac{(B-b)}{b}r\sigma_{ab}^2 + \sigma_a^2\right].$$

Dividing by the degrees of freedom, $(a-1)$, we find

$$E(A^*) = \sigma_e^2 + \frac{(B-b)}{B}r\sigma_{ab}^2 + br\sigma_a^2,$$

as given in Table 2

REFERENCES

[1] Anderson, R L., and Bancroft, T. A , *Statistical Theory in Research.* New York: McGraw-Hill Book Company, 1952

[2] Bennett, Carl A , and Franklin, Norman L., *Statistical Analysis in Chemistry and the Chemical Industry* New York: John Wiley and Sons, 1954 (pp. 470–7).

[3] Cornfield, Jerome, "On samples from finite populations," *Journal of the American Statistical Association*, 39 (1944), 236–9

[4] Cornfield, Jerome, "The two-way classification in the analysis of variance from the point of view of urn sampling," Mimeographed notes, 1953

[5] Eisenhart, Churchill, "The assumptions underlying the analysis of variance," *Biometrics*, 3 (1947), 1–21.

[6] Hald, A , *Statistical Theory With Engineering Application.* New York. John Wiley and Sons, 1952.

[7] Kempthorne, Oscar, *Design and Analysis of Experiments* New York: John Wiley and Sons, 1952.

[8] Kempthorne, Oscar, "The randomization theory of experimental inference," paper delivered to American Statistical Association and Biometrics Society, 1952. In press

[9] Mentzer, Eldo. G., "Tests by the analysis of variance," *WADC Technical Report* 53-23 January, 1953

[10] Mood, Alexander McFarlane, *Introduction to the Theory of Statistics* New York: McGraw-Hill Book Company, 1950

[11] Neyman, J , with the cooperation of K. Iwaskiewicz and St Kolodzieczyk, "Statistical problems in agricultural experimentation," *Journal of the Royal Statistical Society Supplement*, 2 (1935), 107–54.

[12] Pitman, E. J G , "Significance tests which may be applied to samples from any population. III. The analysis of variance test," *Biometrika*, 29 (1937), 322–35.

[13] Scheffé, Henry, "Statistical methods for evaluation of several sets of constants and several sources of variability," *Chemical Engineering Progress*, 50 (1954), 200–5.

[14] Tukey, J. W., "Interaction in a row by column design," Memorandum Report 18. Princeton University, 1949.

[15] Tukey, John W., "Causation, regression and path analysis," in *Statistics and Mathematics in Biology*, Chapter 3, Ames, Iowa: State College Press, 1954.

[16] Vaurio, V. W , and Daniel, Cuthbert, "Evaluation of several sets of constants and several sources of variability," *Chemical Engineering Progress*, 50 (1954), 81–6.

[17] Villars, Donald Statler, *Statistical Design and Analysis of Experiments for Development Research* Dubuque, Iowa. Brown, 1951

[18] Welch, B. L., "On the z-test in randomized blocks and Latin squares," *Biometrika*, 29 (1937), 21–52.

[19] Wilk, M B , "The randomization analysis of a generalized randomized block design," Unpublished M S Thesis. Iowa State College, 1953

[20] Wilk, M. B , "The randomization analysis of a generalized randomized block design," *Biometrika*, 42 (1955), 70–9

[21] Wilk, M B , "The logical derivation of linear models for experimental situations and their use in selecting the appropriate error term in the analysis of variance, I, AV-3 (1953)"; II, "On the analysis of a general randomized block type design, AV-4 (1954)"; III, (with O. Kempthorne) "The expectations of mean squares for a general three-way classification, AV-7 (1954)"; IV, (with O Kempthorne) "Fixed, mixed and random models, AV-8 (1954)", V, "Expectations of mean squares for a general three-way classification (Part 2), AV-9 (1954)"; VI, "Interactions with experimental units, AV-10 (1954) " Analysis of variance project reports, Statistical Laboratory, Iowa State College

COMPONENTS OF A DIFFERENCE BETWEEN TWO RATES*

EVELYN M. KITAGAWA
University of Chicago and Scripps Foundation

WHEN comparing the incidence of some phenomenon in two or more groups, social researchers place much emphasis on the need for holding constant those related factors that would tend to distort the comparison. For example, before comparing the death rates for the residents of two areas, demographers frequently control the factors of differences between the areas in age, sex and race composition. A technique commonly used to accomplish this is "standardization" of the rates for the two areas by relating them both to a standard population with specified age-sex-race composition. By applying the schedule of age-sex-race specific death rates for each of the groups to the age-sex-race composition of the standard population, then noting the total death rate that results, it is possible to compare the death rates for the areas with reasonable confidence that differences in age, sex and race composition do not explain the differences between the rates for the areas that still remain after they have been standardized. Controlling the effect of related factors by this method is termed direct standardization.[1]

It is often noted that such standardized rates are "artificial." For example, an age-sex-race-standardized death rate indicates what the total (or crude)[2] death rate of a population would be *if* it had the age-sex-race composition of the standard population while retaining its own age-sex-race-specific death rates. While such a measure may not be very useful for descriptive purposes, it is an important analytical device.

Since the crude rate of any population is its "real" or "observed" rate in the (descriptive) sense that it is the total rate which results from the particular composition and specific rates which prevail in that population, a systematic statement of relationships between crude and standardized rates for two or more groups may help to bridge the gap

* Expanded version of a paper read at the annual meeting of the American Statistical Association held in Chicago, December 27–30, 1952. The preparation of this manuscript was sponsored jointly by the Population Research and Training Center, University of Chicago, and Scripps Foundation, Miami University—the latter through funds provided by the Rockefeller Foundation for the study of population distribution. The author is indebted to Philip M. Hauser, Donald J. Bogue, O. Dudley Duncan, Beverly Duncan and J. J. Feldman for a careful reading of the paper and many suggestive comments and criticisms.

[1] For a description of the standardization procedure, the assumptions involved, and some of the limitations see A. J. Jaffe, *Handbook of Statistical Methods for Demographers* (Washington: U. S. Govt Printing Office, 1951), Chapter III. For handling cases where a complete schedule of specific rates is not available for the particular areas being compared, but a cross-tabulation of the population by the variables is available, an alternative procedure termed "indirect standardization" has been developed.

[2] The terms "crude," "total," and "unstandardized" are used interchangeably in this paper.

between the "observed" crude (or total) rates and the "artificial" standardized rates. As yet, very little attention has been directed to the problem of formalizing the analysis of standardized rates, and of systematically explaining which factors account for the differences between standardized rates in comparison with corresponding differences between their unstandardized rates. If standardization alters a difference between two total rates, it should be possible to measure the amount of change, and to break it up into components attributable to the various factors for which the data were standardized. Formalizing the process of making inferences from standardized data, and establishing a technique whereby the change accomplished by standardization can be interpreted in terms of the factors involved, are the objectives of this paper.

The technique presented here is called "components of a difference between two rates." It is a revision and refinement of a mode of analysis utilized at the University of Chicago since 1948.[3] The purpose of the technique is to explain the difference between the total rates of two groups in terms of differences in their specific rates and differences in their composition. Thus, the components framework is broader in scope than that of standardized rates, since the framework of the latter is designed to summarize and compare differences in two (or more) sets of specific rates.

The basic concept of separation into components has been used in research for some time. Implicitly, it has been used whenever the size and direction of the difference between standardized rates for two populations was compared with the size and direction of the difference between their crude rates, and the inference made that the "difference between these two differences" is the result of the different compositions of the two populations. Explicitly, it has been used in research under various headings. For example, in the 1920's Ogburn determined what part of the difference between the per cent of the U. S. population married in 1890 and the per cent married in 1920 was due to changes in the age composition of the population between these two dates.[4] The chapter on standardization in Jaffe's *Handbook of Statistical Methods for Demographers* includes a section titled "Removing the influence of changing occurrence rates" which discusses a similar procedure. In his

[3] Earlier statements were prepared by Ralph H. Turner and the writer, with the counsel of Philip M. Hauser. Evelyn Kitagawa, "A Method of Analysing the Influence of Several Non-Quantitative Factors on a Result," (Dept. of Sociology, University of Chicago, March, 1948, hectographed); Ralph Turner, "Whites and Negroes in the Labor Force" (unpublished Ph D. dissertation, University of Chicago, September, 1948); Turner, "The Expected Cases Method Applied to the Nonwhite Male Labor Force," *American Journal of Sociology*, LV (September, 1949), 145–56.

[4] E. R. Groves and W. F. Ogburn, *American Marriage and Family Relationships* (New York: Henry Holt & Co., 1928), 160–2.

example, the difference between the proportion of the total population 5 to 20 years of age attending school in 1940 and the proportion attending school in 1910 is, in effect, separated into two parts, one called the "influence of changes in age distribution" and the other the "influence of changes in occurrence rates."[5] A similar procedure is applied to an analysis of the 1890–1930 change in the proportion of persons gainfully occupied, in an article by Wolfbein and Jaffe.[6]

Edwin C. Goldfield's "method of multiple standardization with allocation of interactions" is concerned with the same general problem but carries the analysis further to include consideration of several factors simultaneously, and to evaluate the net influence of each of the factors as well as their interaction. This method is used, and very briefly described for particular examples, in Durand's study of the labor force.[7]

The revised components framework described in the present article was first used in a study of labor mobility.[8] Any comparison of this definition of components with previous definitions requires a statement of the rationale underlying the components analysis, as well as some standard terminology and algebraic notation. These will be developed in the course of presenting the writer's revised approach, and comparison of the various definitions will be made later.

As has already been noted, components are closely related to standardized rates. Because the method of standardizing rates is familiar to research workers, and because components may be defined by subtracting a standardized rate difference from a crude rate difference, the components analysis will be approached by putting standardized rates in a components framework

The analysis will be developed first with two factors (I and J) controlled. Formulas for one factor (I) will be given later, and an abbreviated extension to control three or more factors will also be discussed. The factors included in the analysis may be either quantitative or non-quantitative.

---

[5] Jaffe, *op. cit.* pp. 44–6. While Jaffe's discussion of the problem emphasizes its interpretation as a method of "holding constant changes in the occurrence rates," in contrast to the conventional method of computing standardized rates where composition is held constant, there is a clear separation into the two parts (or components) mentioned above, in Table 4, p. 46.

[6] S. L. Wolfbein and A. J Jaffe, "Demographic Factors in Labor Force Growth" *American Sociological Review*, XI (August, 1946), 393–6

[7] John D. Durand, *The Labor Force in the United States, 1890–1960* (New York Social Science Research Council, 1948), Appendix B Goldfield's method is very similar to the components analysis described in the 1948 manuscripts of Kitagawa and Turner.

[8] Evelyn M. Kitagawa, "The Relative Importance—and Independence—of Selected Factors in Job Mobility, Six Cities, 1940–49" (Chicago Community Inventory, University of Chicago, hectographed report, 1953).

CONVENTIONAL STANDARDIZATION IN A COMPONENTS FRAMEWORK

Suppose we have observed a difference between the crude rates of two groups, p and P.[9] Also, suppose that data for each group are cross-classified by two factors, I and J.

Let

n_{ij} = number of persons in both the ith category of I and the jth category of J in population p

N_{ij} = number of persons in both the ith category of I and the jth category of J in population P

t_{ij} = rate for persons in the ith category of I and the jth category of J in population p

T_{ij} = rate for persons in the ith category of I and the jth category of J in population P

Also

$n_{.}$ and $N_{.}$ = total number of persons in populations p and P, respectively

$t_{.}$ and $T_{.}$ = rate for total persons in populations p and P, respectively (i e , crude rates of p and P).

The IJ-composition of p and P is designated by $n_{ij}/n_{.}$ and $N_{ij}/N_{.}$ respectively. That is, IJ-composition is a proportionate distribution in which the number of persons in each IJ cell has been divided by the total number of persons in the group In conventional summation notation

$$n_{.} = \sum_i \sum_j n_{ij} \quad \textit{and} \quad N_{.} = \sum_i \sum_j N_{ij}$$

$$t_{.} = \sum_i \sum_j \frac{n_{ij}}{n_{.}} \quad \textit{and} \quad T_{.} = \sum_i \sum_j T_{ij} \frac{N_{ij}}{N_{.}} \text{ by definition.}^{10}$$

If we let p represent the group with the higher crude rate, the differ-

---

[9] The term crude rate is used here to refer to the over-all unstandardized rate of any specified group—that is, we may be concerned with an analysis of the components of the difference between crude rates of two age groups Although such rates are usually called age-specific rates, in this context they are total or crude rates if the two age groups are the population groups with which we are concerned

The concept of a rate in the components analysis includes percentages and means Medians cannot be used because they do not have the algebraic properties of means and percentages which are utilized in the components method.

[10] Since the crude rate of any population may be regarded as the result of its own IJ-specific rates weighted by its own IJ-composition.

ence, $t_{..} - T_{..}$, will be positive. Expressing this difference in the equation

$$t_{..} - T_{..} = \sum_i \sum_j t_{ij} \frac{n_{ij}}{n_{..}} - \sum_i \sum_j T_{ij} \frac{N_{ij}}{N_{..}}$$

makes explicit the fact that the difference between the crude rates of p and P is due to differences in both their IJ-specific rates and their IJ-composition.

As has already been mentioned, conventional standardization techniques can be utilized to compute IJ-standardized rates for p and P, which will summarize differences in their IJ-specific rates *holding constant* differences in their IJ-composition. In the notation outlined above, the difference between IJ-standardized rates for p and P may be expressed as follows:

$$\sum_i \sum_j \frac{n_{ij}}{n_{..}} (t_{ij} - T_{ij}) \text{ if group } p \text{ is used as standard}^{11}$$

$$\sum_i \sum_j \frac{N_{ij}}{N_{..}} (t_{ij} - T_{ij}) \text{ if group } P \text{ is used as standard}$$

$$\sum_i \sum_j \frac{n'_{ij}}{n'_{..}} (t_{ij} - T_{ij}) \text{ if a third group, } p' \text{ is used as standard}$$

Thus, the difference between the standardized rates of two groups is a weighted average of differences in their IJ-specific rates, with the IJ-composition of the standard population used as the weights.

The objective of the components framework is to allocate the difference between two crude rates into components which reflect differences in specific rates of the two groups, on the one hand, and differences in their composition, on the other hand. The equations above suggest that the difference between the IJ-standardized rates of two groups might be used as the "component due to differences in specific rates," since this difference may be considered a measure of their differences in specific rates. Furthermore, if the difference between two standardized rates is subtracted from the corresponding difference between two crude rates, it is easily demonstrated that the result is a weighted average of differences between the composition of the two groups. For example, when the difference between the IJ-standardized rates of groups p and P, with group p used as the standard, is subtracted from

[11] Since $\sum_i \sum_j t_{ij}(n_{ij}/n_{..}) = t_{..}$ = both the crude and IJ-standardized rate for p, and since $\sum_i \sum_j T_{ij}(n_{ij}/n_{..})$ = the IJ-standardized rate for P, when p is used as the standard population.

the difference between their crude rates, the result is

$$(t_{..} - T_{..}) - \sum_i \sum_j \frac{n_{ij}}{n_{..}} (t_{ij} - T_{ij})$$

$$= \sum_i \sum_j t_{ij} \frac{n_{ij}}{n_{..}} - \sum_i \sum_j T_{ij} \frac{N_{ij}}{N_{..}} - \sum_i \sum_j t_{ij} \frac{n_{ij}}{n_{.}} + \sum_i \sum_j T_{ij} \frac{n_{ij}}{n_{..}}$$

$$= \sum_i \sum_j T_{ij} \left(\frac{n_{ij}}{n_{..}} - \frac{N_{ij}}{N_{.}}\right)$$ = weighted average of differences in IJ-composition of p and P, with the IJ-specific rates of P as the weights.

In this case, the difference between the crude rates of p and P may be expressed as the sum of two major components as follows:

$$t_{.} - T_{..} = \sum_i \sum_j \frac{n_{ij}}{n_{.}} (t_{ij} - T_{ij}) + \sum_i \sum_j T_{ij} \left(\frac{n_{ij}}{n} - \frac{N_{ij}}{N_{.}}\right)$$

where

$$\sum_i \sum_j \frac{n_{ij}}{n} (t_{ij} - T_{ij})$$ = component due to differences in IJ-specific rates
= difference between IJ-standardized rates (p as standard)

$$\sum_i \sum_j T_{ij} \left(\frac{n_{ij}}{n} - \frac{N_{ij}}{N_{..}}\right)$$ = component due to differences in IJ-composition

However, these equations show that different standard populations are used as weights in the two components. That is, while p may have been purposely selected as the standard population to provide weights to measure the difference in IJ-standardized rates, the net result—from the components framework—is that the other group, P, is the standard population which provides weights for summarizing differences in IJ-composition of p and P.

Similar results are obtained if the second population, P, is selected as the standard for computing IJ-standardized rates for p and P. Thus, it is also true that

$$t_{..} - T_{..} = \sum_i \sum_j \frac{N_{ij}}{N_{.}} (t_{ij} - T_{ij}) + \sum_i \sum_j t_{ij} \left(\frac{n_{ij}}{n_{.}} - \frac{N_{ij}}{N_{..}}\right)$$

where

$$\sum_i \sum_j \frac{N_{ij}}{N.} (t_{ij} - T_{ij}) = \text{component due to difference in } IJ\text{-specific rates}$$

$$= \text{difference between } IJ\text{-standardized rates } (P \text{ as standard})$$

$$\sum_i \sum_j t_{ij} \left(\frac{n_{ij}}{n} - \frac{N_{ij}}{N} \right) = \text{component due to differences in } IJ\text{-composition.}$$

In this case, when the IJ-composition of P is selected to weight differences in IJ-specific rates, the IJ-specific rates of p are implied as weights for the component due to differences in IJ-composition

When a third population, p', is used as the standard for computing IJ-standardized rates for p and P, the following components are the result:

$$(t. - T.) = \sum_i \sum_j \frac{n'_{ij}}{n'.} (t_{ij} - T_{ij}) + \left[\sum_i \sum_j t_{ij} \left(\frac{n_{ij}}{n} - \frac{n'_{ij}}{n'} \right) + \sum_i \sum_j T_{ij} \left(\frac{n'_{ij}}{n'} - \frac{N_{ij}}{N.} \right) \right]$$

where

$$\sum_i \sum_j \frac{n'_{ij}}{n'..} (t_{ij} - T_{ij}) = \text{component due to differences in } IJ\text{-specific rates of } p \text{ and } P$$

$$= \text{difference between } IJ\text{-standardized rates of } p \text{ and } P \ (p' \text{ as standard})$$

and

$$\sum_i \sum_j t_{ij} \left(\frac{n_{ij}}{n.} - \frac{n'_{ij}}{n'} \right) + \sum_i \sum_j T_{ij} \left(\frac{n'_{ij}}{n'} - \frac{N_{ij}}{N} \right)$$

$$= \text{component due to differences in } IJ\text{-composition of } p \text{ and } P.$$

The last equation makes explicit the weights implied in the "composition component" when a third population, p', is used as the standard for computing standardized rates Specifically, the difference in IJ-

composition between p and P is broken into two parts—the difference between p and p' or

$$\left(\frac{n_{ij}}{n.} - \frac{n'_{ij}}{n'}\right)$$

and the difference between p' and P or

$$\left(\frac{n'_{ij}}{n'} - \frac{N_{ij}}{N.}\right);$$

and IJ-specific rates of p are implied as weights for the first part, while IJ-specific rates of P are the weights for the second part

Thus, the statement that a population, p', is used as a standard population for computing IJ-standardized rates for p and P means specifically that the IJ-composition of p' is used to weight differences in IJ-specific rates of p and P. If, in interpreting the results of standardization, the difference between the standardized rate difference and the crude rate difference is attributed to the different composition of the two groups, it should be recognized that the IJ-specific rates of the two groups themselves are the weights, or standard, for summarizing differences in their IJ-compositions

MAJOR COMPONENTS (TWO FACTORS, I AND J)

The components analysis starts directly and explicitly from the perspective of allocating a crude rate difference into parts attributable to differences in composition and specific rates Suppose the crude rates refer to the two groups, p and P Also, suppose that their specific rates and composition are classified by two factors, I and J. Then, an unambiguous allocation of the crude rate difference into two major components—one reflecting differences in IJ-composition only, and the other differences in IJ-specific rates only—will be obtained if two sets of weights, w_{ij} and w_{ij}', are selected which satisfy the equation

$$t \quad - T\ . = \sum_i \sum_j w_{ij}\left(\frac{n_{ij}}{n.} - \frac{N_{ij}}{N}\right) + \sum_i \sum_j w'_{ij}(t_{ij} - T_{ij})$$

The first component on the right side of the equation represents a weighted sum of differences in IJ-composition, and the second component is a weighted sum of differences in IJ-specific rates. These components will be called "Combined IJ" and "Residual IJ," respec-

tively.[18] A meaningful interpretation of the weights is obtained if the w_{ij} are a set of IJ-specific rates, and the w'_{ij} define an IJ-composition. In this case, the crude rate difference (which is due to differences in both IJ-composition and IJ-specific rates) is expressed as the sum of two components: (1) Combined IJ, or differences between the IJ-composition of p and P, with IJ-specific rates held constant; and (2) Residual IJ, or differences between the IJ-specific rates of p and P, with IJ-composition held constant.

Thus, the chief distinction between the components perspective and that of conventional standardized rates is in the specification of a standard population. In the components framework, the standard population must include a set of IJ-specific rates to be used as weights for the composition component, as well as an IJ-composition to be used as weights for the "specific rates" component.

The results of the preceding section indicate that the IJ-specific rates and the IJ-composition of the *same population* will not satisfy the requirements of a standard population for the components framework *if* the objective is an unambiguous two-component solution.[19] For example, when the IJ-composition of p was used as the standard for computing standardized rates, the IJ-specific rates of P were implied as the weights for the component due to differences in composition. Therefore the three sets of components discussed in the preceding section are excluded as possible two-component solutions unless the standard population is conceived as a population with the IJ-composition of one group and the IJ-specific rates of another group; for example, the composition of p and the specific rates of P. Although for particular problems a researcher may be willing to define such a standard population, the two-component solution proposed below would seem to be more generally useful from the components perspective.

Since the standard population for a two-component solution inevitably involves some characteristic of both p and P, the possibility of using the average composition and the average specific rates of these two groups to define the standard population is suggested In our notation

---

[18] The term "Residual IJ" is assigned to the component due to differences in IJ-specific rates, since it measures the difference between the total rates of the two groups *after* IJ-composition is held constant, while the crude rate difference measures the difference between total rates of the two groups *without* taking into account differences in IJ-composition.

[19] This statement is, of course, limited to the three populations discussed in the previous section as possible standards for computing standardized rates—namely, p, P or a third "observed" population, p'. We have not yet discussed the use of certain "hypothetical" populations. Also, we have thus far required a set of components to allocate the crude rate difference into two parts only, one due to differences in specific rates and the other to differences in composition.

$$\frac{t_{ij} + T_{ij}}{2} = \text{average } IJ\text{-specific rates of } p \text{ and } P$$

$$\frac{\frac{n_{ij}}{n..} + \frac{N_{ij}}{N..}}{2} = \text{average } IJ\text{-composition of } p \text{ and } P$$

A little algebraic manipulation will show that

$$t. \; - T.. = \text{Combined } IJ + \text{Residual } IJ$$

where

$$\text{Combined } IJ = \sum_i \sum_j \frac{t_{ij} + T_{ij}}{2}\left(\frac{n_{ij}}{n} - \frac{N_{ij}}{N.}\right)$$

= component due to differences in IJ-composition (with average IJ-specific rates of p and P as weights)

$$\text{Residual } IJ = \sum_i \sum_j \frac{\frac{n_{ij}}{n\ .} + \frac{N_{ij}}{N.}}{2} (t_{ij} - T_{ij})$$

= component due to differences in IJ-specific rates (with average IJ-composition of p and P as weights)

= difference between IJ-standardized rates of p and P (with average IJ-composition of p and P as the standard)

Thus, a standard population having the average IJ-composition and the average IJ-specific rates of p and P does allocate the difference between their crude rates into two major components, one reflecting differences in their IJ-composition and the other differences in their IJ-specific rates [14]

[14] It may be noted that the use of a weighted average will not yield the symmetric results of the simple average For example, if the specific rates of p are assigned a weight of 2 and those of P a weight of 1 to obtain weighted average specific rates as weights for the "composition component," then in the weights for the "rates component" the composition of P will have a weight of 2 and the composition of p a weight of 1 That is

$$\textit{If} \text{ Combined } IJ = \sum_i \sum_j \frac{2t_{ij} + T_{ij}}{3}\left(\frac{n_{ij}}{n\ .} - \frac{N_{ij}}{N\ .}\right).$$

$$\textit{Then} \text{ Residual } IJ = \sum_i \sum_j \frac{\frac{n_{ij}}{n..} + 2\frac{N_{ij}}{N..}}{3} (t_{ij} - T_{ij}).$$

This solution is proposed here as the most meaningful one for general purposes when the components framework is used. However, an alternative three-component solution may be considered, especially for certain types of comparisons. For example, if the two crude rates refer to the same population at two different dates, the following questions might be asked: (1) How much change would there have been in the crude rate between the two dates if the IJ-composition of the population changed as it did but the IJ-specific rates had remained constant (as of the earlier date)? (2) How much change would there have been in the crude rate if the IJ-specific rates changed as they did, but the IJ-composition had remained constant (as of the earlier date)? (3) If the changes measured in (1) and (2) do not add to the total change in the crude rate, by how much do they fail to do so? If we let p represent the population at the later date, and P the population at the earlier date, the total change in the crude rate between the two dates may be expressed as the sum of three components, as follows:

$$t - T = \sum_i \sum_j T_{ij}\left(\frac{n_{ij}}{n} - \frac{N_{ij}}{N}\right) + \sum_i \sum_j \frac{N_{ij}}{N}(t_{ij} - T_{ij}) + \sum_i \sum_j (t_{ij} - T_{ij})\left(\frac{n_{ij}}{n.} - \frac{N_{ij}}{N}\right)$$

The first (or composition) component measures changes in IJ-composition assuming no change in IJ-specific rates; the second (or rates) component measures changes in IJ-specific rates assuming no change in IJ-composition; and the third (or interaction) component involves changes in both IJ-composition and IJ-specific rates [16]

An exchange of comments with several readers of this paper revealed some differences in preference for the two- and three-component solutions, particularly when the two crude rates refer to the same population at different dates (or to any comparison where one set of events may be considered to precede another) In such a comparison, it may be argued, a three-component solution, with the initial population providing the weights for both the "rates" and "composition" components, appears to be a logical approach. The interpretation of components in this case is described above

On the other hand, it is the writer's opinion that a good case can be

[16] It may be noted that these three components reduce to a two-component solution in either of two ways. If the first and third components are added, the result is a composition component with the specific rates of p as weights. If the second and third components are added, the result is a rates component with the composition of p as weights

made for the two-component solution in such situations. A brief discussion of the reasons for this preference may clarify the rationale of both solutions. First, the selection of a standard population for a components analysis of the change in a crude rate between two dates can be made from any of several assumptions The three-component solution derives a rates component assuming no change in composition, a composition component assuming no change in specific rates, and an interaction component reflecting changes in both rates and composition. However, changes in rates and composition are seldom independent—rather, a change in one is likely to affect the other. It may be argued, therefore, that since both were changing during the period, a logical set of weights for summarizing changes in specific rates, for example, would be the average composition of the population during the period; similarly, the weights for the composition component might be the average specific rates experienced by the population throughout the period The two-component solution uses averages for the two dates as weights for these components. While such averages are not equivalent to averages for regular intervals throughout the period, they are often the only averages which can be obtained, also, even when data are available, the computation of annual averages, or averages for any other regular interval, would be a very laborious task It may be noted in this connection that the simple average for the beginning and end of the period will equal the annual (or other regular interval) average if it is assumed that changes were distributed uniformly throughout the period.

Second, the selection of a standard population for a components analysis of crude rates for two dates is in many respects comparable to the problem of selecting weights for index numbers. Economists give careful consideration to alternative sets of weights when computing index numbers for two dates, and an average of appropriate values for the two years is frequently used [16] For example, to compute an index of the physical volume of manufacturing production for 1947 relative to 1939, the Bureau of the Census defined change in physical volume as the "change in value of net ouput, or value added [by manufacture], at constant prices," and used the average prices for the two years as weights to be applied to the quantities of each product included in the index.[17] An advantage of the use of average prices, as compared with prices for the first year, is that it avoids overweighting products whose

[16] The average used in Fisher's "ideal" index is the geometric mean, while the simple arithmetic mean is used in the Marshall-Edgeworth formula.

[17] Bureau of the Census, *Census of Manufactures 1947, Indexes of Production*, pp. 2–4

prices have shown a considerable "relative" decrease between the two years, and also avoids underweighting products whose prices have shown a considerable "relative" increase during the period.

In the general case—when the crude rates refer to two different groups, neither of which may be considered to precede the other—there is no "initial state" from which to measure change and, therefore, less emphasis on the desirability of using the rates and composition of one group as weights for both components. An unambiguous allocation into two components, using the average rates and average composition as weights, is a possible solution, and appears more logical than the various alternatives considered.[18]

SUBCOMPONENTS OF COMBINED IJ

Combined IJ, the composition component of the two-component solution, may be further divided into three subcomponents as follows:[19]

Combined IJ = Net I_J + Net J_I + Joint IJ *where*

$$\text{Net } I_J = \sum_i \sum_j \left[\left(\frac{t_{ij} + T_{ij}}{2} \right) \left(\frac{\frac{n_{\cdot j}}{n_{\cdot\cdot}} + \frac{N_{\cdot j}}{N_{\cdot\cdot}}}{2} \right) \right] \left(\frac{n_{ij}}{n_{\cdot j}} - \frac{N_{ij}}{N_{\cdot j}} \right)$$

$$\text{Net } J_I = \sum_i \sum_j \left[\left(\frac{t_{ij} + T_{ij}}{2} \right) \left(\frac{\frac{n_{i\cdot}}{n_{\cdot\cdot}} + \frac{N_{i\cdot}}{N_{\cdot\cdot}}}{2} \right) \right] \left(\frac{n_{ij}}{n_{i\cdot}} - \frac{N_{ij}}{N_{i\cdot}} \right)$$

$$\text{Joint } IJ = \sum_i \sum_j \left[\frac{t_{ij} + T_{ij}}{2} \right] \frac{\left(\frac{N_{ij}}{N_{i\cdot}} \frac{n_{i\cdot}}{n_{\cdot\cdot}} - \frac{n_{ij}}{n_{i\cdot}} \frac{N_{i\cdot}}{N_{\cdot\cdot}} \right) + \left(\frac{N_{ij}}{N_{\cdot j}} \frac{n_{\cdot j}}{n_{\cdot\cdot}} - \frac{n_{ij}}{n_{\cdot j}} \frac{N_{\cdot j}}{N_{\cdot\cdot}} \right)}{2}$$

[18] If it is desired, a range of variation for each component, as a result of the different weights which might be used in its computation, might be estimated by computing each component twice—once with one of the two populations providing the weights, and a second time with the other population providing the weights. Thus, the two values for each component defined by the two equations

$$t_{\cdot\cdot} - T_{\cdot\cdot} = \sum_i \sum_j T_{ij} \left(\frac{n_{ij}}{n_{\cdot\cdot}} - \frac{N_{ij}}{N_{\cdot\cdot}} \right) + \sum_i \sum_j \frac{n_{ij}}{n_{\cdot\cdot}} (t_{ij} - T_{ij})$$

and

$$t_{\cdot\cdot} - T_{\cdot\cdot} = \sum_i \sum_j t_{ij} \left(\frac{n_{ij}}{n_{\cdot\cdot}} - \frac{N_{ij}}{N_{\cdot\cdot}} \right) + \sum_i \sum_j \frac{N_{ij}}{N_{\cdot\cdot}} (t_{ij} - T_{ij})$$

might be used to estimate the range of variation for each component. It may be noted that each major component in the two-component solution proposed in this paper is an average of the two values obtained for this component from these equations.

In these formulas,[20] $n_{i.}$ and $N_{i.}$ represent the total number of persons in the ith category of factor I in groups p and P, respectively; and $n_{.j}$ and $N_{.j}$ represent the total number of persons in the jth category of factor J in p and P, respectively.

Net I_J is a weighted sum of differences in I_J-composition (I-composition *within* the J subgroups)[21] of p and P and, therefore, may be interpreted as that part of the difference between their crude rates which is attributable to differences in net I_J-composition, or I-composition *independent* of J. The average IJ-specific rates and average gross J-composition of p and P are used as weights in this subcomponent.[22] That is, Net I_J measures differences in I_J-composition applied to a standard population which has the average IJ-specific rates and the average gross J-composition of p and P.

Similarly Net J_I is a weighted sum of differences in J_I-composition (J-composition *within* the I subgroups) and may be interpreted as that part of the crude rate difference which is due to differences in net J_I-composition, or J-composition *independent* of I. The average IJ-specific rates and average gross I-composition of p and P are used as weights for this subcomponent. That is, Net J_I measures differences in J_I-composition applied to a standard population which has the average IJ-specific rates and the average gross I-composition of p and P.

Joint IJ is that part of the Combined IJ component which cannot be allocated to differences in net I_J-composition or to differences in net J_I-composition. It represents the part of the crude rate difference which is accounted for by differences in combined IJ-composition but which cannot be allocated independently to I or J. Its equation shows it to be a weighted sum of differences in IJ-composition which result from combining the net composition of one of the two populations with the gross composition of the other,[23] using the average IJ-specific rates of the two populations as weights.

---

[19] Because the composition component of the three-component solution is not a total measure of differences in IJ-composition—part of the composition differences are included in the interaction component—there is no discussion of its subcomponents.

[20] Note that $n_{i.} = \Sigma_j n_{ij}$, $n_{.j} = \Sigma_i n_{ij}$, etc.

[21] In the equation for this subcomponent, $(n_{ij}/n_{.j} - N_{ij}/N_{.j})$ represents the differences in I-composition *within* the subgroups of factor J.

[22] The weights are enclosed in brackets, [ ], in the equations defining the subcomponents. Gross J-composition refers to the per cent distribution of a group when it is classified by factor J only. For example, in group p gross J-composition is defined by $n_{.j}/n_{..}$

[23] Specifically, the term

$$\frac{N_{ij}}{N_{.j}}\frac{n_{i.}}{n_{..}} - \frac{n_{ij}}{n_{.j}}\frac{N_{i.}}{N_{..}}$$

expresses the difference between (1) the net J_I-composition of P combined with the gross I-composition of p, and (2) the net J_I-composition of p combined with the gross I-composition of P. Similarly, the term

Thus, a complete components analysis allocates the difference between the crude rates of p and P into four parts:

$$t_{..} - T_{..} = \text{Net } I_J + \text{Net } J_I + \text{Joint } IJ + \text{Residual } IJ$$

where Net I_J is the part due to differences in I-composition independent of J; Net J_I is due to differences in J-composition independent of I; Joint IJ is due to differences in Joint IJ composition, or to differences in Combined IJ composition which cannot be allocated independently to I or J; and Residual IJ is due to differences in IJ-specific rates of p and P. The standard population used for this purpose is one having the average IJ-composition and the average IJ-specific rates of p and P.

ONE FACTOR COMPONENTS

If data for the two groups, p and P, are cross-classified by only one factor, I, a two-component allocation of the difference between their crude rates may be defined as follows:

$$t_{.} - T_{.} = \text{Gross } I + \text{Residual } I$$

where

$$\text{Gross } I = \sum_i \frac{t_i + T_i}{2}\left(\frac{n_i}{n} - \frac{N_i}{N}\right)$$

$$\text{Residual } I = \sum \frac{\dfrac{n_i}{n} + \dfrac{N_i}{N}}{2}(t_i - T_i).$$

The Gross I component represents that part of the difference between the crude rates of p and P which is due to differences in their I-composition, and Residual I the part due to differences in their I-specific rates Residual I is also equal to the difference between the

---

$$\frac{N_{ij}}{N_{.j}}\frac{n_{.j}}{n} - \frac{n_{ij}}{n_{.j}}\frac{N_{.j}}{N}.$$

expresses the difference between (1) the net I_J-composition of P combined with the gross J-composition of p, and (2) the net I_J-composition of p combined with the gross J-composition of P These two sets of differences—one for the combination of net I_J and gross J compositions, the other for the combination of net J_I and gross I compositions—are obtained because the two factors I and J, are interchangeable in this subcomponent. Therefore, corresponding differences (that is, differences for the same IJ cell of the cross-classification) in these two sets are averaged and the resulting set of average differences are weighted by average IJ-specific rates of p and P. This interpretation of the Joint IJ subcomponent is presented to state explicitly the differences which are measured. It is not necessary to compute these differences to obtain this subcomponent since it may be obtained more simply by subtracting the sum of subcomponents Net I_J and Net J_I from the major component Combined IJ

I-standardized rates of p and P, with the average I-composition of p and P as the standard population.

A three-component allocation similar to that described for two factors, is as follows.

$$t - T = \sum_i T_i\left(\frac{n_i}{n} - \frac{N_i}{N}\right) + \sum_i \frac{N_i}{N}(t_i - T_i) + \sum_i (t_i - T_i)\left(\frac{n_i}{n} - \frac{N_i}{N}\right)$$

In this case, the I-specific rates and I-composition of group P are used as weights for the "composition" and "rates" components, respectively The third component is an "interaction" component which is due to differences in both composition and rates.

INTERPRETATION OF RESULTS

We shall first consider the two-component solution The interpretation of the two major components and the subcomponents is relatively simple when all the components are positive, since the difference in crude rates can then be attributed partly to differences in IJ-specific rates and partly to differences in IJ-composition, I_J-composition, J_I-composition and Joint IJ-composition In this situation, the components may be converted to per cent components, with the crude rate difference as the base. For example,

$$\frac{\text{Combined } IJ}{t - T} \times 100 = \text{per cent of difference between crude rates which is due to differences in } IJ\text{-composition of } p \text{ and } P.$$

Even in this case, however, two qualifications should be kept in mind: (1) the components include the effects of hidden forces behind the factors, I and J, and nothing in the components technique justifies inferences as to causal relationships—such inferences must be based on knowledge outside the statistical technique itself; (2) a small Residual component may mask larger influences, in opposite directions, of factors not held constant in the analysis. The latter qualification means, for example, that Residual IJ may be less than Residual IJK; that is, the difference in IJK-standardized rates may be greater than the difference in IJ-standardized rates Thus, the results of a components analysis should be interpreted as "applicable within the frame-

work of the particular factors held constant." For example, a Residual *IJ* per cent component of 35 per cent may correctly be interpreted as indicating that "the difference in *IJ*-standardized rates is only 35 per cent as great as the difference in crude rates," and the corresponding Combined *IJ* component of 65 per cent may be interpreted as indicating that "65 per cent of the crude difference is attributable to *IJ*-composition." But, as additional factors (*K*, *L*, etc.) are held constant, the Residual component may not always decrease (and the Combined component increase) with the addition of each new factor—in fact, it will not do so unless all the other factors (*K*, *L*, etc.) operate in the same direction. This characteristic simply points to the fact that the difference between two crude rates is not the equivalent of a concept like total variance of a dependent variable in regression analysis, for example, which will be increasingly "explained" as more independent variables are added to the regression equation.[24]

When not all of the components are positive, their interpretation is more complicated. For example, if Residual *IJ*—the difference between *IJ*-standardized rates—is greater than the difference between the crude rates, Combined *IJ* is negative If all of the subcomponents have the same sign (that is, are negative), Residual *IJ* might be used as the base for per cent components, which can be interpreted as follows:

$$\frac{t_{..} - T_{..}}{\text{Residual } IJ} \times 100 =$$ per cent of the difference between *IJ*-standardized rates which is observed or evident in the difference between their crude rates.

$$\frac{-\text{ Combined } IJ}{\text{Residual } IJ} \times 100 =$$ per cent of the difference between *IJ*-standardized rates which is obscured (in the crude rates) by differences in *IJ*-composition. (This is a positive component because Combined *IJ* is negative.)

$$\frac{-\text{ Net } I_J}{\text{Residual } IJ} \times 100 =$$ per cent of the difference between *IJ*-standardized rates obscured by differences in I_J-composition, etc.

Other combinations of positive and negative components do not

[24] Strictly speaking, the addition of a new independent factor need not increase the amount of "explained variance," but it cannot decrease it.

readily lend themselves to meaningful per cent components. In such situations, the components themselves may be used without conversion to per cents. For example, if $t_{..} - T_{..} = 2$, Residual $IJ = -25$, and Combined $IJ = 27$, we might simply say that while the crude rate of p exceeded that of P by 2 points, the standardized rate of P exceeded that of p by 25 points, and that differences in IJ-composition were responsible for these widely different results The Net I_J, Net J_I, and Joint IJ subcomponents may be used to indicate the importance of I-composition independent of J, J-composition independent of I, and Joint IJ-composition in obscuring (in the crude rates) the difference between standardized rates

If a crude rate difference is allocated into three major components, and all are positive, each component might be expressed as a per cent of the crude rate difference. Or, if this solution is used in a comparison over time, it may be desired to use the crude rate of the earlier date as the per cent base That is, we may write these components as

$$t_{..} = T_{..} + \sum_i \sum_j T_{ij}\left(\frac{n_{ij}}{n_{..}} - \frac{N_{ij}}{N_{..}}\right) + \sum_i \sum_j \frac{N_{ij}}{N_{..}}(t_{ij} - T_{ij}) + \sum_i \sum_j (t_{ij} - T_{ij})\left(\frac{n_{ij}}{n_{..}} - \frac{N_{ij}}{N_{..}}\right)$$

If each of the terms in this equation is expressed as a per cent of $T_{..}$, the crude rate at the earlier date, then the per cent which $t_{..}$ is of $T_{..}$ is equal to 100 plus the three percentage components In this case, for example,

$$\frac{\sum_i \sum_j T_{ij}\left(\frac{n_{ij}}{n_{..}} - \frac{N_{ij}}{N_{..}}\right)}{T_{..}} \times 100 =$$ per cent change in the total rate between the two dates as a result of changes in composition, assuming no change in specific rates during the period.

AN EXAMPLE OF RESULTS

In a study of labor mobility in six cities made in 1951,[25] this method was used to determine the extent to which city differences in job mo-

[25] This research, conducted by the Chicago Community Inventory of the University of Chicago, was based on data obtained in the Six-City Mobility Study, one of the industrial manpower research studies sponsored by the United States Air Force under Project SCOOP Findings are summarized in Evelyn M. Kitagawa, "Relative Importance and Independence of Selected Factors in Job Mobility, Six Cities, 1940–49," *op cit*

bility rates were due to differences in the composition of the labor force in the various cities. To cite one particular example, the crude mobility rate (mean number of jobs held, 1940–49) of Los Angeles men was 32 per cent higher than that of Philadelphia men, or an average of 3.14 jobs as compared with 2.37 in Philadelphia.

Table 1 presents data on mobility rates and composition, by migrant status and time spent in the labor force from 1940 to 1949, for men in these two cities. In the lower half of the table are the results of a components analysis of the difference between crude mobility rates of Los Angeles and Philadelphia men with these two factors held constant—migrant status (J) and time spent in labor force (I)

Examination of the specific mobility rates shows that Los Angeles men were consistently more mobile than Philadelphia men, even when data are cross-classified by migrant status and time in the labor force. However, it is also clear that migrants were more mobile than non-migrants, and persons in the labor force 5–9½ years were more mobile than persons in for less or more time Furthermore, the percentage distributions which describe the composition of men in the two cities indicate higher proportions of Los Angeles men in the high mobility categories—for example, 47 per cent of the Los Angeles men were migrants as compared with 13 per cent of the Philadelphia men Thus, we expect that part, but not all, of the difference between crude rates for men in these two cities is due to differences in their composition with respect to migrant status and time spent in the labor force. The components analysis quantifies this relationship.

Differences in composition with respect to both migrant status and time spent in the labor force account for 47 per cent of the difference between crude rates of Los Angeles and Philadelphia men. And, the difference between IJ-standardized rates for men in these two cities is 53 per cent of their crude rate difference, using their average IJ-composition as standard (Residual IJ)

Furthermore, migrant composition alone, independent of time spent in the labor force, accounted for 38 per cent of the crude rate difference, with only 1 per cent due to composition by time spent in the labor force independent of migrant composition, and 7 per cent to these two factors jointly.

Similar components were computed for selected pairs of cities, with these and other factors held constant, to determine which factors were most important in accounting for city differentials in crude mobility rates and to what extent these city differentials reflected differences in specific (or standardized) mobility rates. Space does not permit a more detailed analysis, but the purpose here is only to illustrate the use of the method in a specific set of data.

Measures of the sampling variability of per cent components have not been determined Since each component is based on the difference between two sums of a large number of products, its sampling variance may prove to be too unwieldy to estimate, though further study of this problem might yield some simplifying assumptions which will furnish approximate estimates of sampling variance without overburdening computations [26] With census data based on complete counts or with very large samples where cross-classifications do not run thin, per cent components should be relatively stable or reliable. But with small samples or with small cell frequencies in complete counts, inferences should be made with caution

COMPARISON WITH PREVIOUS DEFINITIONS OF COMPONENTS

Reference was made in the introductory paragraphs to previous work involving the allocation of a crude rate difference into components The set of components proposed in this paper differs from those used previously in two respects. First, the earlier approaches selected one of the two populations being compared to provide weights for one of the two major components, considered this population the standard population for the components analysis, and did not make explicit the weights used in the other major component which was obtained as a residual (by subtracting the computed component from the crude rate difference, or an equivalent procedure) The algebraic presentation in this paper has made explicit the set of weights which was implicit in such a two-component solution; for example, when the IJ-specific rates of one group are used to weight differences in IJ-composition, the IJ-composition of the other group is used to weight differences in IJ-specific rates The two-component solution proposed here uses a standard population having the average IJ-composition and the average IJ-specific rates of the two groups

Second, the rationale of a set of subcomponents in this paper is more defensible than that in the previous literature.[27] This can best be seen by noting that earlier definitions of the Joint IJ subcomponent, if

[26] One place in the analysis where tests of significance can be made is in the determination of whether one population's specific rates are *on the whole* larger than another's. The test consists of considering each sub-group in one population with its comparable sub-group in the other population as a four-fold table One can then sum the chi-squares from the individual four-fold tables (See Karl Pearson and J F Tocher, "On Criteria for the Existence of Differential Death Rates," *Biometrika*, 11 (1916), 159–64, S A Stouffer and Clark Tibbitts, "Tests of Significance in Applying Westergaard's Method of Expected Cases to Sociological Data," *Journal of the American Statistical Association* 28 (1933), 293–302, H F Dorn and S A Stouffer, "Criteria of Differential Mortality," *Journal of the American Statistical Association*, 28 (1933), 402–13).

[27] To the writer's knowledge, the only previous work involving subcomponents is contained in the cited reference to Goldfield's method and the Turner and Kitagawa references. The comments here are applicable, for the most part, to the Turner and Kitagawa definitions, although Goldfield's definition of a Joint subcomponent was similar, he did not retain it as a measure of Joint composition but allocated it back to the factors involved.

TABLE 1

JOB MOBILITY RATES (MEAN NUMBER OF JOBS HELD 1940–49) AND COMPOSITION (PERCENTAGE DISTRIBUTION) BY MIGRANT STATUS AND TIME SPENT IN THE LABOR FORCE, FOR LOS ANGELES AND PHILADELPHIA MEN: 1940–49*

Migrant Status and Time in Labor Force 1940–49	Mobility Rates		Composition (%)		Difference (L A. Minus Phila.)	
	Los Angeles	Phila-delphia	Los Angeles	Phila-delphia	Rates	Compo-sition
All Men	3.14	2 37	100	100	77	0
Less than 5 yrs.	2.90	2 42	11	7	48	4
5 but less than 9½ yrs	3.82	3 26	30	26	56	4
9½–10 yrs.	2.84	2 03	59	67	81	−8
Migrants	3.77	3.13	47	13	.64	34
Less than 5 yrs.	2 89	2 29	6	1	60	5
5 but less than 9½ yrs.	4.07	3 43	17	4	64	13
9½–10 yrs.	3 79	3.15	24	8	64	16
Non-migrants	2 58	2 26	53	87	.32	−34
Less than 5 yrs.	2 92	2.45	5	6	.47	−1
5 but less than 9½ yrs.	3 49	3.23	13	22	.26	−9
9½–10 yrs.	2.20	1 88	35	59	.32	−24

Components of the Difference Between the Crude Job Mobility Rates of Los Angeles Men (3 14) and San Francisco Men (2 37)
(Two Factors. I = time spent in labor force 1940–49; J = migrant status)

Name of Component	Per Cent Component ($t_{..} - T_{..} = 100$)	Actual Component ($t_{..} - T_{..} = .77$)
Combined IJ	47	.359
Net J_I	38	.296
Net I_J	1	.008
Joint IJ	7	054
Residual IJ	53	.411

* Data refer to a probability sample of 1,313 men in Los Angeles and 1,571 men in Philadelphia who worked one month or more in 1950. Persons residing in each city in 1951 who had resided there 11 years or less (that is, who moved there after the beginning of the 1940–49 decade) were classified as migrants for purposes of this study.

translated into the algebraic notation, do not reduce to a weighted sum of differences in composition resulting from a combination of the net composition of one group with the gross composition of the other. In previous definitions the Net I_J and Net J_I subcomponents were defined independently, and then the Joint IJ subcomponent was computed by subtracting Net I_J and Net J_I from Combined IJ, without determining, algebraically, what such a residual measured For example, when the specific rates of P were used as weights for the Combined IJ component in previous definitions, the subcomponents were defined as follows, if translated into our notation:

$$\text{Net } I_J = \sum_i \sum_j T_{ij} \frac{n_{.j}}{n} \left(\frac{n_{ij}}{n_{.j}} - \frac{N_{ij}}{N_{.j}} \right)$$

$$\text{Net } J_I = \sum_i \sum_j T_{ij} \frac{n_{i.}}{n} \left(\frac{n_{ij}}{n_{i.}} - \frac{N_{ij}}{N_{i.}} \right)$$

$$\text{Joint } IJ = \sum_i \sum_j T_{ij} \left(\frac{N_{ij}}{N_{i.}} \frac{n_{i.}}{n} - \frac{N_{ij}}{N} - \frac{n_{ij}}{n} + \frac{N_{ij}}{N_{.j}} \frac{n_{.j}}{n} \right).$$

However, if the rationale of a set of subcomponents in this paper is used to compute subcomponents for the same Combined IJ component

$$\left(\text{Combined } IJ = \sum_i \sum_j T_{ij} \left[\frac{n_{ij}}{n} - \frac{N_{ij}}{N} \right] \right),^{28}$$

the results could be
either

$$\text{Net } I_J = \sum_i \sum_j T_{ij} \frac{n_{.j}}{n} \left(\frac{n_{ij}}{n_{.j}} - \frac{N_{ij}}{N_{.j}} \right)$$

$$\text{Net } J_I = \sum_i \sum_j T_{ij} \frac{N_{i.}}{N} \left(\frac{n_{ij}}{n_{i.}} - \frac{N_{ij}}{N_{i.}} \right)$$

$$\text{Joint } IJ = \sum_i \sum_j T_{ij} \left(\frac{N_{ij}}{N_{.j}} \frac{n_{.j}}{n} - \frac{n_{ij}}{n_{i.}} \frac{N_{i.}}{N} \right)$$

[28] Two similar sets of subcomponents could be obtained if the rates of p (instead of P) were used as weights in the Combined IJ component; equations for these two sets can be written by interchanging the roles of p and P in the weights of the equations above Thus, four sets of subcomponents might be defined for the two Combined IJ components. Averaging the four values for each subcomponent would give a single set of subcomponents which is identical to the set defined in this paper

or

$$\text{Net } I_J = \sum_i \sum_j T_{ij} \frac{N_{.j}}{N_{..}} \left(\frac{n_{ij}}{n_{.j}} - \frac{N_{ij}}{N_{.j}} \right)$$

$$\text{Net } J_I = \sum_i \sum_j T_{ij} \frac{n_{i.}}{n} \left(\frac{n_{ij}}{n_{i.}} - \frac{N_{ij}}{N_{i.}} \right)$$

$$\text{Joint } IJ = \sum_i \sum_j T_{ij} \left(\frac{N_{ij}}{N_{i.}} \frac{n_{i.}}{n} - \frac{n_{ij}}{n_{.j}} \frac{N_{.j}}{N_{..}} \right).$$

COMPARISON WITH WESTERGAARD'S METHOD OF EXPECTED CASES

The logic and mechanics of Westergaard's method of expected cases is described by Woodbury in a study of infant mortality rates.[29] Jaffe describes the method as "essentially an elaboration of the conventional standardization technique which can be used in a situation where the investigator wishes to isolate the influence of a single factor from that of other associated factors."[30]

Basically, the method involves the computing of ratios of the actual number of cases in a particular group to the expected number of cases in the same group should it retain its own composition while being exposed to the specific rates of a standard population (with the total of the various subgroups usually used as standard). For example, Woodbury in the study cited above, uses the method to isolate the influence of order of birth on infant mortality holding constant age of mother and earnings of father. If we let factor I represent age of mother and factor J earnings of father, then in our notation we might say

$n_{ij}^{(K)}$ = number of births in both the ith category of I and the jth category of J in the population of k-order births

N_{ij} = number of births in both the ith category of I and the jth category of J in the population of total births

$t_{ij}^{(K)}$ = infant mortality rates for births in the ith category of I and the jth category of J in the population of k-order births

T_{ij} = infant mortality rates for births in the ith category of I and the jth category of J in the population of total births

His results are summarized in a table similar to Table 2. A comparison of the Westergaard ratios in the second column is used to indicate the extent of the variation in infant mortality by order of birth with the

[29] R. M. Woodbury, "Westergaard's Method of Expected Deaths as Applied to the Study of Infant Mortality," *Journal of American Statistical Association*, 18 (1923), 366–76, and reproduced in Jaffe, *op. cit.*, Chap. III.

[30] Jaffe, *op. cit.*, p. 48.

influence of factors I and J eliminated. Also, comparison of the ratios in the first and second columns (the former measure crude or unadjusted birth-order differentials in infant mortality) indicates the influence of factors I and J—that is, IJ-composition—on birth-order differentials in infant mortality rates However, the Westergaard technique does not quantify this aspect of the analysis—that is, it does not measure what part of the crude birth-order differentials is attributable to factors I and J

If we compare the Westergaard framework with conventional standardization and the components analysis, several relationships become evident First, the components framework applied to the analysis of birth-order differentials in mortality might be used to analyze the difference between crude infant mortality rates of any two birth orders into components due to differences in IJ-composition on one hand and to differences in IJ-specific rates on the other hand That is to say, it quantifies a relationship which is apparent in a Westergaard analysis

TABLE 2

Order of Birth	Ratio of Original Rate to Average Rate (Col 1)	Ratio of Actual to Expected I & J Constant (Col. 2)
Total	$\frac{T_{..}}{T_{..}} = 1\ 00$	$\frac{\sum_i \sum_j T_{ij} N_{ij}}{\sum_i \sum_j T_{ij} N_{ij}} = 1\ 00 \left(= \frac{T}{T} \right)$
First	—	—
—	—	—
—	—	—
kth	$\frac{t_{..}^{(K)}}{T_{..}} = \frac{\sum_i \sum_j t_{ij}^{(K)} \frac{n_{ij}^{(K)}}{n_{..}^{(K)}}}{\sum_i \sum_j T_{ij} \frac{N_{ij}}{N}}$	$\frac{t_{..}^{(K)} n_{..}^{(K)}}{\sum_i \sum_j T_{ij} n_{ij}^{(K)}} = \frac{\sum_i \sum_j t_{ij}^{(K)} n_{ij}^{(K)}}{\sum_i \sum_j T_{ij} n_{ij}^{(K)}}$
		$= \frac{\sum_i \sum_j t_{ij}^{(K)} \frac{n_{ij}^{(K)}}{n_{..}^{(K)}}}{\sum_i \sum_j T_{ij} \frac{n_{ij}^{(K)}}{n_{..}^{(K)}}}$
—	—	—
—	—	—
—	—	—
Tenth	—	—

but not rigorously measured in terms of the "components" concept defined in this paper

To the writer's knowledge, there is no published work comparing the Westergaard technique with conventionally standardized rates selected to accomplish the same purpose, although the technique has always been recognized as a "standardization" method. Examination of the formulas for Westergaard's ratios in the table above reveals that these ratios are actually equivalent to ratios of indirectly standardized infant mortality rates for each birth order (with the population of total births as standard) to the crude infant mortality rate for total births (i.e., the standard population). For example, the indirectly standardized rate for k-order births, with total births as the standard population, is given by[31]

$$\frac{\text{actual infant deaths to } k\text{-order births}}{\text{expected infant deaths to } k\text{-order births}} \times \begin{pmatrix}\text{crude death rate of} \\ \text{standard population}\end{pmatrix}$$

or

$$\frac{\sum_i \sum_j t_{ij}^{(K)} n_{ij}^{(K)}}{\sum_i \sum_j T_{ij} n_{ij}^{(K)}} T_{..}$$

And if this is divided by the crude rate ($T_{..}$) for total births we get the Westergaard ratio.[32]

Thus, if the objective is to compare the relative incidence of mortality from one birth order group to the next, holding constant the disproportionate weighting of the various groups by age of mother and earnings of father, the Westergaard ratios might be considered as approximations in the same sense that indirect standardization approximates the results of direct standardization Ratios based on directly standardized rates,[33] with the population of total births as standard, would be defined by

$$\frac{\sum_i \sum_j t_{ij}^{(K)} \frac{N_{ij}}{N_{..}}}{T_{..}} \quad \text{or} \quad \frac{\sum_i \sum_j t_{ij}^{(K)} \frac{N_{ij}}{N_{..}}}{\sum_i \sum_j T_{ij} \frac{N_{ij}}{N_{..}}} \quad \text{or} \quad \frac{\sum_i \sum_j t_{ij}^{(K)} N_{ij}}{\sum_i \sum_j T_{ij} N_{ij}}.$$

[31] Jaffe, *op. cit.*, pp 44–8.

[32] The Westergaard ratio for a category may also be described as the correction factor which is applied to the crude rate of the standard population to obtain the indirectly standardised rate for that category.

[33] Such ratios have been used in at least one study—Donald J Bogue, *A Methodological Study of Migration and Labor Mobility in Michigan and Ohio in 1947* (Oxford. Scripps Foundation for Research in Population Problems, 1952), p. 64.

The only factor which varies in these ratios for different birth orders are the IJ-specific rates of the birth orders, while in the Westergaard ratios the composition which appears in the numerator and denominator of each ratio also varies from one birth order to the next. That is to say, Westergaard ratios do not strictly speaking, hold IJ-composition constant from one birth order to the next. However, there may be good reasons for computing ratios of indirectly standardized rates such as the Westergaard ratios [34]

EXTENSION OF FRAMEWORK TO THREE VARIABLES

Extending the components framework to three variables (I, J and K) is simple for the major components—Combined IJK and Residual IJK. In this case

$$t_{...} - T_{...} = \text{Combined } IJK + \text{Residual } IJK$$

where

$$\text{Combined } IJK = \sum_i \sum_j \sum_k \left(\frac{t_{ijk} + T_{ijk}}{2}\right)\left(\frac{n_{ijk}}{n_{...}} - \frac{N_{ijk}}{N_{...}}\right)$$

and

$$\text{Residual } IJK = \sum_i \sum_j \sum_k \left[\frac{\dfrac{n_{ijk}}{n} + \dfrac{N_{ijk}}{N}}{2}\right](t_{ijk} - T_{ijk}).$$

However, expressions for net and joint subcomponents would be quite complex if the two-factor model were extended to three factors. A much simpler solution results if one is willing to consider two of the three factors as a single factor and apply the two-factor framework. For example, if we are primarily interested in the influence of I independent of both J and K, we might consider I as one factor and the cross-classification of J by K as a second factor—the latter will be denoted by (JK) Then, in addition to the combined and residual components defined above, we could define the following subcomponents:

$$\text{Net } I_{JK} = \sum_i \sum_j \sum_k \left(\frac{t_{ijk} + T_{ijk}}{2}\right)\left[\frac{\dfrac{n_{.jk}}{n} + \dfrac{N_{.jk}}{N}}{2}\right] \cdot \left(\frac{n_{ijk}}{n_{.jk}} - \frac{N_{ijk}}{N_{.jk}}\right) \quad .$$

[34] Two of the most obvious are (1) when IJ-specific rates are available only for total births and not for each birth order, and (2) when enough of the IJ-specific rates for the various birth orders are based on too small numbers for stability. See Peter Cox, *Demography* (London Cambridge University Press, 1950), Chapter 7, for a statement of weights implicit in indirect standardization.

$$\text{Net } (JK)_I = \sum_i \sum_j \sum_k \left(\frac{t_{ijk} + T_{ijk}}{2}\right) \cdot \left[\frac{\frac{n_i}{n_{..}} + \frac{N_i}{N_{..}}}{2}\right] \left(\frac{n_{ijk}}{n_i} - \frac{N_{ijk}}{N_i}\right)$$

$$\text{Joint } I(JK) = \sum_i \sum_j \sum_k \left(\frac{t_{ijk} + T_{ijk}}{2}\right) \cdot \left[\frac{\frac{N_{ijk}}{N_i}\frac{n_{i..}}{n} - \frac{n_{ijk}}{n_{.jk}}\frac{N_{.jk}}{N} + \frac{N_{ijk}}{N_{.jk}}\frac{n_{.jk}}{n} - \frac{n_{ijk}}{n_{i..}}\frac{N_i}{N_{..}}}{2}\right].$$

In this situation, Net I_{JK} represents the part of the difference between crude rates attributable to differences in I-composition independent of both J and K, while Net $(JK)_I$ measures the part attributable to combined JK-composition independent of I

Such an analysis could be made with any pair of the factors considered as a single factor. Also, if four or more factors are combined in any way to reduce to two, a similar approach could be used.

RECENT DEVELOPMENTS IN THE SAMPLING OF HUMAN POPULATIONS IN GREAT BRITAIN*

C. A. MOSER

Division of Research Techniques, London School of Economics

This article describes the main developments in the sampling of human populations in Great Britain during the last five years Changes in methodology, as well as new applications, are included

A PREVIOUS article, published in this Journal in 1949 [44], described the main British activities in the field of sample surveys of human populations at that time. Since then the picture has changed in various respects and it seems worth while now to bring the earlier account up to date The purposes of this article are to describe the major changes that have taken place in sampling practice, to mention some new applications and to give a guide to the somewhat diffuse literature Once again, only surveys concerned broadly with human populations[1] are discussed, emphasis throughout is on methods rather than results.

DEVELOPMENTS IN METHODOLOGY

Developments in the methods used can be grouped under four heads: sampling frames; types of sample design; refinements in random (probability) sample designs, and research on methodology.

Sampling Frames. A determining feature of sampling practice anywhere is the nature of the sampling frames available In the U.K., random samples for most large-scale surveys are selected from one or another of the national population lists which are at hand Their existence largely explains why area sampling has not been used here.

In 1949 three lists were mentioned as being particularly useful to samplers: the National Register, the Rating List and the Register of Electors. The general view about their relative suitability for sampling was expressed by Gray and Corlett [24]: " . we would advocate the use of the Rating Records for samples of households or housewives and the National Register for samples of individuals, or particular age- and sex-groups, of demobilized persons, and also for surveys of the town-planning type". The paper referred to gives a full account of the various frames.

* I should like to express my thanks to statisticians and other experts in Government departments and elsewhere and to colleagues at the London School of Economics for their comments on a draft of this paper The paper in its final form is entirely my responsibility and the comments made in no way commit the experts who kindly gave me advice

[1] One or two surveys of economic institutions, e.g., firms, are included.

The National Register, however, was only available to official bodies so that most of the organizations in the sample survey field could not use it even if they wanted to. This, among other things, led the Social Survey to conduct an inquiry into the main alternative, the Register of Electors, and a report on the findings was published by Gray, et al., [25]. In some respects this frame always has been attractive to the sampler. It is available centrally and locally, its order and arrangement are convenient, and the individuals on it are numbered serially, thus facilitating quasi-random selection. The misgivings about the Register had been that not everyone entitled to be on it would in fact appear on it and that, since it was revised only once a year, it was always somewhat out of date. More precisely, since the Register is published 4 months after the so-called "qualifying date" for registration, it is that much out of date when it appears and 16 months out of date by the time it is replaced.

The results of the Social Survey's sample check on the accuracy of the Register were reassuring. About 96% of those entitled to be on it were in fact listed, 92% still being at their registered address. The sample check took place 7 months after the qualifying date and the authors estimated that, at the time of replacement of the old Register, 87% of those entitled to be on it would still be at their registered address. The effect of the initial loss of 4% on the composition of samples was not, in the authors' view, likely to be serious. More important were the removals, which accounted for the remaining inaccuracy. Movers tend to contain a high proportion of young people so that there is a danger of bias (cf., the findings in a recent readership survey [28]) Gray, et al., discussed this danger and proposed a method for dealing with the problem in practice

Broadly speaking, it now seemed that the Register of Electors could with suitable care be used as a frame for samples of individuals and households. The importance of these findings was enhanced by the discontinuation of the National Register in 1952. Since then the Register of Electors has become the most commonly used frame for random samples, both by the Social Survey and others. Edwards [19] has described the practical problems involved in its use. The Rating Records are the only other frame commonly used for large-scale surveys; several inquiries based on it will be mentioned below.

Types of Sample Design. The author's 1949 article referred to the occasional practice—in otherwise random sample designs—of selecting first-stage units, say towns, purposively. It also discussed the continuingly wide use of quota sampling and commented that this method "should . . . not be employed in any surveys on which administrative

decisions are based." Now, in 1955, the use of purposive selection in the above manner and that of quota sampling have virtually disappeared from such surveys

In market and opinion research, the economic attractions of quota sampling are still regarded by many as sufficient to commend its use, but even here a tendency increasingly to use random methods is discernible.

Refinements in Random Sample Designs.[2] To begin with, we should note the influence upon random sampling practice in this country of the work of the Government Social Survey. Most of the designs now used in official and other surveys bear some resemblance to those initiated by the Survey and described in the paper by Gray and Corlett [24].

(a) *Stratification Factors.* In most national surveys of the general population administrative districts (county boroughs, urban districts and rural districts, etc) or parliamentary constituencies are used as first-stage units, when there are more than two sampling stages, wards or polling districts are the most common second-stage units, addresses, households or individuals being the final-stage units The first-stage units are usually stratified geographically and by urban/rural district. Administrative districts, which vary enormously in size, used also to be stratified by population size At the time of the previous article, these stratifications were the only ones in general use, intermediate-stage units, such as wards or polling districts, were generally not stratified at all, because of the lack of suitable data The main weakness in all this was that neither the typical first-stage nor second-stage units were generally stratified by anything related to economic or social characteristics. Since then, two indexes relevant for this purpose have been developed. Both are due to the Social Survey.

The industrialization index. This index, which is described by Gray and Corlett [24], is the "ratio of industrial to total rateable value in each district" and appears to be a useful index of the economic character of a district With its development, the Social Survey turned to a new type of sampling design· administrative districts are no longer stratified by size of population, but by industrialization index (as well as region and urban/rural district), the selection within strata then being made with probability proportionate to size This type of design has been widely followed by other organizations

The J-index. The more serious weakness of earlier designs was the

[2] Two qualifications must be made about this section. In the first place, it will apply only to samples of the general population; there is no space to discuss refinements introduced in the many surveys of special groups Second, theoretical developments are not discussed.

lack of data by which either wards or polling districts could be stratified. Indeed, it was because of this that many random sample designs, including those of the Social Survey, were designed on a two-stage basis, with no stage intermediate between administrative districts (or constituencies) and the final-stage units. This, of course, was costly in terms of field-work The Social Survey's search for possible stratifications resulted in their development of the so-called "J-index", described by Gray, Corlett and Jones [26]. On the Register of Electors, electors liable for jury service are marked with the letter "J". As qualification for jury service is a function of the rateable value of the premises occupied by the elector, it seemed reasonable that the "percentage of registered electors with this suffix in any area might serve as an indicator of the economic status of the area" Tests of the index confirmed this. The index has the further advantage that it can readily be calculated for any area, whether ward, polling district or, for that matter, administrative district, constituency or any arbitrary area. It is used regularly by the Social Survey and occasionally by other organizations. The set of J-indexes for England and Wales is expected to be published in the near future. Corlett has given some additional details on the index in [11] and [12].

Another factor for stratification which is occasionally used is the proportion of voters voting, say, Conservative, at the last election. If constituencies are the first-stage units, this is a good stratification to use. (Constituencies have several points to commend them For one thing, they do not vary as much in size as administrative districts; for another, the Register of Electors is arranged by constituencies) Other possible stratification factors might be derived from the 1951 Census of Population. for instance, the proportion of occupied persons in an area who finished their full-time education at a given age; the proportion of adults in an area in a given social class (the criterion being occupation), housing density and so forth.

Possibilities of stratifying households, addresses or individuals are limited It would be practicable to stratify a Register of Electors' sample by sex, as the sex of most individuals can be deduced from their Christian names More useful would be stratifications by age, income or occupation but none of these are feasible with our national sampling frames (except, of course, after selection)

(b) *Other Points*. Selection with varying probabilities has been a regular feature of population sampling in the U.S. for many years and, as already mentioned, has become fairly general practice here. Edwards [19] describes a national four-stage random sample in which,

at the first stage, administrative districts were selected from within strata with probability proportionate to population, then polling districts were selected within the chosen administrative districts again with probability proportionate to population; the third-stage units were households and the fourth-stage units individuals. Over-all probabilities were kept equal The problem of whether to sample with or without replacement has been tackled in various ways Sometimes the difficulty has been by-passed by selecting systematically within strata (see the example in [24]), at others the selection has been made without replacement, but the sampling errors calculated (much more simply) as if it had been with replacement. Very occasionally, the selection has been made with replacement, the sampling units being included more than once if necessary This is inconvenient, though not as serious a drawback as is sometimes suggested; in the official Household Expenditure Enquiry described below, one rural district was included twice In this connection we may refer to a result in a paper by Durbin [15] to the effect that, if sampling is done without replacement but the standard error calculated as if it had been with replacement, the true standard error will be overestimated, but by only a small amount

As a result of the new stratification indexes described above, national samples are now usually three-stage designs, where formerly two-stage designs were more common Although the resultant saving in cost is obvious, surprisingly little is known, or at any rate published, about the effects of different degrees of clustering on variances and costs, about the benefits of the various stratifications or, for that matter, about the precision attained in general. As the users of survey results come to attach greater importance to the size of sampling errors, so samples will be designed more with an eye to easing the sampling error calculations If this point is not borne in mind during the design stages, these computations can become prohibitively complicated. Practitioners are then tempted to make rough "back-of-the-envelope" estimates of sampling errors by using the formulas applying to simple random sampling Since the samples are generally stratified multi-stage designs, these formulas are inappropriate and misleading. They may underestimate the true sampling error by a considerable margin (see Lydall [33])

It often happens that the population to be covered differs from that of the most suitable frame. Such was the case in the survey described by Edwards [19] to which reference has already been made. Here a sample of individuals aged 16 and over had to be drawn from the

Register of Electors, which covers adults aged 21 and over. The solution adopted was to use the Register for drawing a sample of households and then to select individuals from a list of household members drawn up by the interviewer The procedure used was that suggested by Kish [30] and was found to work well.

Research on Methodology One of the important changes since 1949 is the growing attention given to research into survey methodology. To some extent every survey organization conducts research into its own methods but it is a fact that most of them are too busy with commissioned surveys to devote much time and staff to research. Some valuable findings of course come about as by-products of everyday surveys but, unless the surveys have been designed with this in view, the results usually lack the decisiveness of experimental conclusions.

Outside the survey organizations themselves, one of the developments to note is the establishment, at the London School of Economics, of the Division of Research Techniques, which devotes a part of its resources to research on survey methods The Division has been financed since its inception in 1949 by the Nuffield Foundation and is directed by M. G Kendall. It is charged with the task of studying "those branches of methodology required in the research undertaken by the School", so that its attention to survey methods is logical.

Four of its researches should be mentioned although, since all have been published, there is no need to go into detail. The first experiment set out to test differences in performance and results of experienced professional and inexperienced student interviewers. Though the former were of course expected to be more successful in obtaining interviews, one did not know how big the differences would be or what effects it would have on results. The evidence on these two questions is presented in papers by Durbin and Stuart [18] and Booker and David [4]

The Division's second field research program was into the nature and efficiency of quota sampling A report on the practice of quota sampling in this country—published by Moser [43]—served as the preliminary stage The central part of the research was an experiment designed to study the relative accuracy and precision of random sampling and of different types of quota sampling. The findings have been published by Moser and Stuart [45]. The discussion following the paper, and published with it, contains the views on the quota sampling

controversy of a number of experts in this country. The Division's third experiment, which dealt with variability in the coding process, has been described by Durbin and Stuart [16].

The Division's most recent field experiment was designed to study the question of call-backs in random sample surveys and, at the same time, the effects of different degrees of clustering on variances and costs. The results were published by Durbin and Stuart [17]; they included some reassuring findings on interviewer variability. A theoretical paper on the non-response problem has been published by Durbin [14] The Division, like many other bodies interested in sample surveys, is now turning its attention increasingly to the study of response errors; it is currently engaged in research on interviewer variability (see also [42]) and is planning a project on memory errors.

Two general features of the Division's survey researches may be of interest. In the first place, they all are based on experimental designs in the modern statistical sense. Secondly, they are planned by a Committee on which members of the Division are joined by representatives of the Social Survey, the BBC Audience Research Department and several of the leading market and opinion research organizations The fieldwork is carried out by the professional organizations and the planning and execution of the projects is a joint effort between the academic members and survey practitioners. Experience during the last few years suggests that this cooperation is fruitful for both sides.

The Division is not the only academic institution in this country with an interest in survey methods, although there seem to be hardly any others which devote a substantial part of their resources to research into survey methodology per se. The Department of Applied Economics of Cambridge University has, in the course of its research on social accounting, done a good deal of work on the methodology of family expenditure studies and on the use of sampling in examining national or local accounts Among publications reporting on the former are papers by Utting and Cole [55] and [56]; among those dealing with the latter, and therefore with samples of documents and institutions, are those of Utting [54] and Stone, et. al , [51] The Oxford Institute of Statistics has produced many points of methodological value in the course of its Savings Survey (see, in particular, Lydall [33])

NEW APPLICATIONS

We shall not attempt anything like a comprehensive coverage of new

applications since 1949. The few cases selected are intended to illustrate the gradual widening of the field to which sampling methods are being applied

Official Sample Surveys

1. *The Government Social Survey*

The Social Survey's role in government administration has not altered since 1949 Gray and Corlett [24] gave an illustrative list of surveys carried out between 1946–49; a later description of the Survey's work is given by Moss [46]. Of methodological interest have been several surveys concerned with the prediction of behavior, such as the enquiry into the demand for campaign medals [59] At the end of the war some 7 million people were entitled to wear medals and, if all of them were to claim their medals, this would have meant striking about 20 million medals. The Survey was asked to predict the likely demand and used an attitude-scale type of approach to do so Their prediction that only about a third of those entitled to medals would apply for them proved very close to the actual demand; as a result some £200,000 was saved.

An earlier survey predicting the effects of freeing the distribution of domestic coal was useful in its demonstration that, at that stage, such a measure would have had disastrous effects on the export surplus In a very different field, a recent study of the use of telephone directories [31] enabled the Government to effect various economies and again save a substantial sum of money

A study of considerable methodological interest, especially in the analysis of data, is the attempt to predict the success of training at Borstal institutions and in Approved Schools The results and methods are described in a book by Mannheim and Wilkins which will be published shortly by H M.S O. A short technical paper by Wilkins [58] has already appeared

A current inquiry into hospital nursing is employing sampling techniques for timing the observations made in a number of hospital wards. These observations will be used in conjunction with attitude studies amongst patients and nurses to gauge the effects of proposed changes in the organization of nursing duties

On the subject of health the Survey has done an interesting study of General Practice. Two articles by Gray and Cartwright have dealt respectively with "Choosing and Changing Doctors" and with the issue and use of medicines [22, 23]

A considerable part of the Survey's work is concerned with consumer expenditure. Over the last few years there have been inquiries into expenditure on laundry, dry cleaning and shoe repairs; household textiles and furnishing fabrics; house repairs and domestic service; meals in restaurants; pharmaceutical products, holidays, hairdressing and cosmetics; and other sectors of expenditure. These surveys are conducted for the Central Statistical Office to assist in its estimates of national expenditure.

Another field in which the Survey is particularly active is nutrition The National Food Survey is mentioned below, there have also been inquiries into the diets of young children and of other special groups

We may also mention at this point a large-scale anthropometric survey carried out for the Clothing Industry Development Council in 1951–1952 (see [8]), since a member of the Social Survey staff—W. F. F Kemsley—was responsible for the analysis and report The purpose of the survey was to assist manufacturers of ready-made clothing with the problem of sizing In order to secure the necessary data, 37 measurements were taken on each of 5000 women between the ages of 18 and 70 and results on these measurements were then given for women of different ages and occupations, with different numbers of children and so forth More important, the relationships between the various measurements have been established.

These are only a few of the varied activities of this organization. Far more of its work—both in the way of survey reports and of methodological research—is now published than was the case in 1949; indeed, this is obvious from the references already given Altogether the Survey's work stands at a high level of technical efficiency, and it is a tribute to its reputation and to the growing appreciation of the role of sample surveys that it has been able to continue through these years of economic pressure. It is to be hoped that this appreciation will remain and grow and that the Survey's potentialities will be fully exploited by successive administrations

We should also note and welcome the increasing extent to which the Social Survey has been brought into the field on the borderlines of social and economic research. It has been linked with a part of the social mobility enquiry at the London School of Economics, with the Savings Survey at Oxford, with the work of the Department of Applied Economics at Cambridge (in this case to give technical advice on matters of design and analysis) and with a current survey among university graduates by Political and Economic Planning These are only a few of many instances.

2. *The General Register Office*

One of the features remarked on in 1949 was the failure of the General Register Office to avail itself of sampling methods in its demographic work and, in particular, in the Population Census. The hope was then expressed that the Registrars-General would decide to use sampling in the 1951 Census. In the event, sampling was not used in the collection of data; but, for the first time, a preliminary analysis was based on a sample of the Census schedules, allowing some results to be published within 18 months of the Census data [20]. Details of the sampling procedure are given in the report, together with a discussion of the representativeness of the sample and some adjustments made to the results An interesting discussion of the accuracy of the sample appears in Yates [60], section 10 16.

The speed with which the preliminary Census results became available amply justified the Registrars'-General decision to analyze a sample of the schedules in the first instance It is to be hoped that they will decide in favor of using sampling in the *collection* of, at any rate, some of the data at the next Population Census Some questions, of course, must be asked of everyone, but on others sufficient precision could probably be attained with a sample coverage. The use of multiphase sampling in the U S Population Census could well serve as an example There is also a strong case for preceding each Census with a series of sample surveys designed to test various forms of questions, definitions, methods of enumeration and so forth; and for following each Census by checks on the quality of data obtained, rather on the lines of the Post-Enumeration Surveys conducted by the U. S. Bureau of the Census.

3. *The Ministry of Food*

The National Food Survey, described in the author's previous article, still continues, but with its coverage now extended from the working-class to the whole population. Four reports on the methods and results of the Survey have been published [35, 36, 37, 38] and others are promised. As from 1953, the fieldwork, coding and analysis have been taken over by the Social Survey.

4. *The Ministry of Labour Household Expenditure Enquiry*

The present U.K Index of Retail Prices is based on weights broadly in accordance with the results of the 1937/8 Ministry of Labour Family Expenditure Survey [41]; actually they were revised somewhat

as from January 1952 to represent the estimated 1950 pattern of consumption. Popular feeling, as well as technical opinion, had for some time favored bringing the weights properly up to date as soon as conditions permitted and a report of the Cost of Living Advisory Committee in 1952 [40] proposed that preparations should be made for a new Household Expenditure Enquiry. The proposal was accepted by the Government and the inquiry began in January 1953, continuing for one year. In determining the sampling plan, interviewing and other methods, experimental inquiries conducted by the Social Survey over many months were of very great value. The inquiry itself was directed by the Ministry of Labour, in consultation with the Central Statistical Office. In Northern Ireland the field work was carried out by the Northern Ireland Ministry of Labour, in Great Britain it was shared between the Ministry of Labour and the Social Survey. Few details about the methods of the inquiry have as yet been published and only some general points can be made here.

The inquiry covered the entire (non-institutional) population of the country, although some classes of households may be excluded when computing the weights for the new index The Government, however, accepted the advice of the Cost of Living Advisory Committee that the additional information obtained from covering the entire population would be of great value and would add proportionately little to the over-all cost

Each sampled household was asked to provide expenditure data for a period of 3 consecutive weeks and the sample (a two-stage design) was so arranged that, for each geographical region as well as for the country as a whole, it was spread evenly over the year This solution of the seasonal problem seems preferable to the more usual one of trying to get households to cooperate for, say, four weeks at quarterly intervals. There were many features of methodological interest in this inquiry, for instance in the steps taken to secure cooperation, in the method of collecting the data (a mixture of interviewing and record-keeping) and in the schedules themselves; however, discussion of these must await publication from an official source.

5. *Ministry of Labour Age-Analysis of Employed Persons*

The Ministry of Labour makes annual estimates of the total number of employees in the country from the quarterly exchange of employment insurance cards (see note in [44]). As the insurance cards do not state date of birth, these analyses do not give separate figures for age groups—except for a division into under 18 and 18 and over (these

groups have different contribution rates). The dates of birth are, however, recorded on the ledger accounts kept by the Ministry of National Insurance and these serve as the sampling frame for the special analysis which produces detailed figures for each industry by age and sex, regional age distribution and data on the relation of age and unemployment; the analysis is described in [39].

6 *Ministry of Education*

An interesting case of a sample of a specialized population was provided in a survey designed "to obtain an estimate of the mean score on a certain reading test of pupils aged 15 in grant-aided schools in England and Wales". The sample was a two-stage one, with Local Education Authorities' areas as first-stage units and schools as second-stage units. All the pupils of the required age in the selected schools were tested. The areas were stratified into urban and rural, and areas were then selected with probability proportionate to population. The schools were stratified by five types and selection was again made with probability proportionate to size. The sample design and estimation of sampling errors is discussed by Peaker [49]

7. *Ministry of Agriculture*

The author's previous article described the sampling scheme of the National Farm Survey A similar type of design has been employed in the inquiries into earnings and conditions of employment in agriculture conducted by the Ministry of Agriculture. This was described by Palca and Davies [48].

The purpose is stated to be as follows: "In the agricultural industry. there is a statutory determination of minimum rates of wages related to a standard working week, and overtime rates and certain other elements of farm workers' remuneration are also statutorily determined . . . there was so far no information on average weekly earnings" and it is this gap the surveys seek to fill. They cover hours of work and numbers employed by occupation as well as earnings

8. *Commissioners of Inland Revenue*

A weak link in British economic statistics is the relatively scanty information about the distribution of personal incomes. Although ingenious estimates have been made from time to time, no straightforward official statistics have been available.

The source from which estimates normally begin are the annual

reports of the Commissioners of Inland Revenue, the authority responsible for the administration and collection of income tax Generally these reports—which are based on personal income tax returns—give distributions of incomes from specific sources (e.g, salaries, profits, rent) not distributions of income aggregated for each person. However, for incomes assessed in the years 1949–50, the Commissioners took a 10% sample of income tax schedules, aggregated the incomes of various kinds on each schedule and produced income distributions. The main results are given in two reports [9, 10] Of particular interest are the breakdowns, in the former report, of income by household composition and those in the second, by geographical region

Though useful, these figures suffer from the drawback that, since a married man's income tax schedule is likely to include his wife's earnings, the derived distributions refer neither precisely to the incomes of individuals nor to those of families. In addition, there must always be some doubt about the accuracy of any figures derived from personal income tax returns

9. *Board of Trade Censuses of Production and Distribution*

An important development was the decision to use sampling in the 1953 Census of Production (i e , the one covering production in the year 1952) In order to reduce the burden upon industry, it was decided to ask fewer questions generally and to require returns from only a sample of firms All firms over a given size in each trade were asked to make returns, but only a sample of the smaller firms Thus, most firms were in effect excused from the necessity of making returns.

This made it possible to publish early estimates of the most important aggregates like total sales, materials used, stocks and capital expenditure. According to the statement announcing the decision to use sampling [1], a sample coverage was feasible only because detailed figures of output and material had been obtained for 1951 and were not required again for 1952. Such details could not have been satisfactorily obtained by sampling The 1954 Census (for production in 1953) is being conducted on a similar sample basis.

The Government has now announced [3] that a full Census of Production is to be taken for the years 1957 and 1961 but that, even on these occasions, detailed returns will be required only from firms employing 25 persons and more. A Sample Survey of Production is to be taken in each year when no full Census is conducted.

The same statement announces that a Census of Distribution is to

be taken once every ten years and that, during the interval, Sample Surveys of Distribution should be made every third or fourth year. This summarises very briefly the official policy regarding the future of the Censuses of Production and Distribution. It represents a major step forward in the role of sampling for the collection of official statistical data.

10. *Board of Trade Capital Expenditure Surveys*

In 1950 and 1951 the Board of Trade conducted inquiries "to ascertain the level of capital expenditure by manufacturing industry in Great Britain" [2]. There had been a Census of Production in respect of 1948, which covered capital expenditure, but these special inquiries made results available more rapidly and enabled capital expenditure to be related to calendar years.

Public Opinion and Market Research

There are no fundamental changes to report in this field, although one can discern a distinct raising of methodological standards. This is due both to increasing competition and to the spread of technical knowledge; that practitioners are taking an increasing interest in methodological advances is beyond question.

In the field of public opinion research, there are no new applications of note known to the writer, but one may point to the success of the British Institute of Public Opinion in forecasting the general elections of the last few years, as well as local by-elections. Kendall and Stuart [29] describe a method for predicting the distribution of parliamentary seats from that of votes in the country. A paper by Durant [13] discusses some of the problems of the Gallup Poll.

The BBC's activities in studying radio listening habits and tastes were described in the author's previous article. The only development to note is the natural progression from "Listener Research" to "Audience Research." Attempts to cover the television audience meet with the inevitable difficulty of small numbers; the problems of research in this field are discussed by Silvey [50].

A field of activity which has become increasingly important in recent years is that of newspaper readership surveys. The interest of their results to newspaper proprietors and advertisers does not need stressing and there are now every year two or three sets of results available from different sources. We may refer to a booklet published by the Market Research Society [34] which gives a history of readership surveys, explains their problems and makes suggestions for future work. In 1954

readership surveys were published by the Hulton Press [27] and by the Institute of Practitioners in Advertising [28].

Two other market research activities, slightly outside the normal run of consumer surveys, may be mentioned. One is the Consumer Panel, noted in the previous article, run by Attwood Statistics Ltd Wadsworth [57] has discussed the Panel from the point of view of the user. The other activity which seems to gain importance is the retail panel or audit in which information derives from retailers rather than from consumers. The work of A C. Nielsen Ltd in this field is well-known and there is another retail audit panel in this country, run by the British Market Research Bureau, the working of which has been described by Treasure [53]

One market research casualty of recent years was the British Export Trade Research Organization, a body devoted to overseas market research. A discussion of the reasons for its failure was published by Treasure [52]

We have remarked on the increasing interest of market research practitioners in methodological advances Indicative of this growing willingness to come out into the open, discuss their methods with each other and with academic experts, and to participate in methodological research, is the foundation of the Market Research Society and the good attendance at its meetings and weekend schools.[8] Even more indicative is the increasing willingness to publish Whereas the author's previous article was unable to instance a single major paper published on market research in this country, quite a few can now be mentioned. To the ones already noted may be added the paper by Cauter [7] describing in detail the organization and work of a large-scale market research agency.

Social and Economic Research

There remain the numerous sample surveys conducted by universities, research institutes and individuals Year by year this type of research activity grows, partly, of course, as a consequence of the development of sampling; many current projects are only practicable because of sampling. There is an attendant danger, in so far as research workers are sometimes led to attempt inquiries on too extensive a scale and to spread their resources too thinly. However, this can be prevented if statistical advice is taken, and, by and large, the influence of sampling is much to the good.

[8] The Association of Incorporated Statisticians held a weekend school on sample surveys late in 1953 Several of the talks given on this occasion are to be published in a collection of papers on British Market Research edited by J Downham, E. Shankleman and J Treasure

In the author's previous article a number of sample surveys in this field were described and many of these continue, some in extended form. The Population Investigation Committee's inquiry into maternity (see [44]) has grown into a "National Survey of the Health and Development of Children" A large number of the children covered in the original (1946) sample are still included in the investigation and records of their physical development, their progress in school, their reading ability and intelligence are being kept As a long-term study this is unique in social research. The National Foundation for Educational Research in England and Wales is joining the P.I.C. in undertaking a national survey of standards of attainment in reading and arithmetic at various ages

In the educational field we can also mention the inquiry—undertaken by Political and Economic Planning and the Social Survey—into the jobs currently held by 1950 university graduates and particularly into the use of arts and science graduates in industry. A survey into the social and educational background of applicants for places in universities in 1955 and 1956 is in its early stages. This is sponsored by the Committee of Vice-Chancellors and Principals A research team at the London University Institute of Education is conducting a national sample survey of the school teaching profession

Findings of the research on social mobility, which has been continuing at the London School of Economics for some years (see [44]), has recently been published in a volume edited by Glass [21]. Much of the material in this book is based upon a survey into the social origin, education and occupational achievement of a national sample of some 10,000 adults. Ancillary investigations included work on the social prestige of occupations, the relation of education and social mobility, child upbringing and social class, the relation of social mobility and marriage, recruitment in a number of professions and the structure of voluntary organizations

Surveys on the travel habits of the people of London and Bristol have been undertaken by the London Transport Executive [32] and the British Transport Commission [6] The Social Survey is currently making a major study of travel in London for the Road Research Laboratory and London Transport Executive. This attempts to get a measure of the effects of the cost of transport delays as well as to provide information on the demand and use of various transport services

There continues to be much research into the expenditure and saving patterns of different population groups. We have already mentioned the official 1953 inquiry and the work of the Oxford Institute of Sta-

tistics and the Cambridge Department of Applied Economics in this field.

The National Health Service and the standard of General Practice have been the subject of a certain amount of research. The British Medical Association sponsored an inquiry into General Practice [5], one part of which was a field study of a sample of 188 General Practices, the other a postal inquiry directed to the remaining National Health Service general practitioner principals

These are only a handful of many current or recent social and economic research inquiries The reader wishing to get an idea of the immense range of activity in this field should consult the register published annually by the National Institute of Social and Economic Research [47]

CONCLUDING REMARKS

The four or five years since the time of the previous article have seen considerable advances in the sampling of human populations in this country The field of application has widened, scepticism of the reliability of sampling seems to have become less widespread, and the standard of methodology has been raised This applies almost equally to official, commercial and academic sample surveys

The causes of these changes are manifold Among them certainly is the spread of knowledge of how sampling works, which itself comes from the increasing attention given to the subject in universities and from the growing literature There are now available several good books on the theory and practice of sample surveys and university courses on the subject are attended by large numbers of students. Among the causes, too, is the interest newly taken by the universities in methodological research in this field Most important of all is the growing willingness of those who conduct—and those who commission—surveys to publish their methods and results This must inevitably serve to stimulate interest and to raise the standard of efficiency

REFERENCES

[1] Board of Trade, "Announcement on census of production," *Board of Trade Journal*, 161 (1951), 1326.

[2] Board of Trade, "Capital expenditure by manufacturing industry in Great Britain," *Board of Trade Journal*, 160 (1951), 827–30 and 162 (1952), 4–6.

[3] Board of Trade, "Report of the committee on the censuses of production and distribution," London H.M S O., Cmd 9276, (1954).

[4] Booker, H. S , and David, S T , "Differences in results obtained by experienced and inexperienced interviewers," *Journal of the Royal Statistical Society*, A, 115 (1952), 232–57.

[5] British Medical Association, "Survey of general practice, 1951–52," *British Medical Journal* 2 (1953), 683–706.

[6] British Transport Commission, "Bristol on the move," *British Transport Commission*, (1953)

[7] Cauter, T., "Organisation of a statistical department: some details of a market research agency," *The Incorporated Statistician*, 3 (1952), 37–53.

[8] Clothing Industry Development Council, "Third Annual report," London: H.M.S.O., (1953).

[9] Commissioners of Inland Revenue, *94th Report for the Year Ended*. 31.3.1951, London: H M.S.O., Cmd. 8436, (1952).

[10] Commissioners of Inland Revenue, *95th Report for the Year Ended* 31 3.1952, London: H.M.S.O, Cmd. 8726, (1953).

[11] Corlett, T., "A use for the jury qualification in sample design," *Applied Statistics*, 1 (1952), 34–6.

[12] Corlett, T., "A change in the proportion of jurors," *Applied Statistics*, 2 (1953), 193–5.

[13] Durant, H., "The Gallup Poll and some of its problems," *The Incorporated Statistician*, 5 (1954), 101–10.

[14] Durbin, J., "Non-response and call-backs in surveys," *Bulletin of the International Statistical Institute*, 34 (1954), 72–86.

[15] Durbin, J., "Some results in sampling theory when the units are selected with unequal probabilities," *Journal of the Royal Statistical Society*, B, 15 (1953), 262–9.

[16] Durbin, J., and Stuart, A, "An experimental comparison between coders," *The Journal of Marketing*, 19 (1954), 54–66

[17] Durbin, J. and Stuart, A, "Callbacks and clustering in sample surveys: an experimental study," *Journal of the Royal Statistical Society*, A, 117 (1954), 387–428.

[18] Durbin, J., and Stuart, A, "Differences in response rates of experienced and inexperienced interviewers," *Journal of the Royal Statistical Society*, A, 114 (1951), 163–205.

[19] Edwards, F., "Aspects of random sampling for a commercial survey," *The Incorporated Statistician*, 4 (1953), 9–25.

[20] General Register Office, "Census 1951, Great Britain, one per cent sample tables," Pts. I and II, London H.M.S.O. (1952).

[21] Glass, D. V (Ed.), "Social mobility in Great Britain," London: Routledge and Kegan Paul (1954)

[22] Gray, P G., and Cartwright, A, "Choosing and changing doctors," *The Lancet*, 265 (1953), 1308–9

[23] Gray, P. G, and Cartwright, A., "Who gets the medicine?" *Applied Statistics*, 3 (1954), 19–28

[24] Gray, P. G., and Corlett, T., "Sampling for the Social Survey," *Journal of the Royal Statistical Society*, A, 113 (1950), 150–206.

[25] Gray, P. G., Corlett, T., and Frankland, P., "The register of electors as a sampling frame," London: Central Office of Information, (1950).

[26] Gray, P. G., Corlett, T., and Jones, P., "The proportion of jurors as an index of the economic status of a district," *The Social Survey* (1951)

[27] Hulton Press, *The Hulton Readership Survey, 1954*, London: 1954.

[28] Institute of Practitioners in Advertising, *The National Readership Survey—1954 Edition*, London: 1954

[29] Kendall, M. G., and Stuart, A., "The law of the cubic proportion in election results," *British Journal of Sociology*, 1 (1950), 183–96.

[30] Kish, L., "A procedure for objective respondent selection within the household," *Journal of the American Statistical Association*, 44 (1949), 380–7.

[31] Lamberth, D. L., "A forecast of effects of economies in London Telephone Directory Services," *The Social Survey*, 1952.

[32] London Transport Executive, "London travel survey, 1949," *London Transport Executive*, 1950

[33] Lydall, H. F., "The methods of the Savings Survey," *Bulletin of the Oxford Institute of Statistics*, 16 (1954), 197–244

[34] Market Research Society, *Readership Surveys A Comparative Study.* London: 1954

[35] Ministry of Food, *Domestic Food Consumption and Expenditure, 1950,* London: H.M.S.O., 1952.

[36] Ministry of Food, *Domestic Food Consumption and Expenditure, 1951.* London H.M.S.O , 1954.

[37] Ministry of Food, *Domestic Food Consumption and Expenditure, 1952,* London· H M.S.O , 1955.

[38] Ministry of Food, *The Urban Working-class Household Diet, 1940–1949.* London: H.M.S O., 1951.

[39] Ministry of Labour, "Age analysis of employed persons," *Ministry of Labour Gazette*, 59 (June 1951), 223–9

[40] Ministry of Labour, *Interim Report of the Cost of Living Advisory Committee.* London: H M S.O , Cmd 8328, 1951

[41] Ministry of Labour, "Weekly expenditure of working-class households in the U.K., 1937–38" *Ministry of Labour Gazette*, 48, 49 (Dec 1940, January and February, 1941)

[42] Moser, C. A., "Interview bias," *Revue de l'Institut International de Statistique*, 19 (1951), 28–40.

[43] Moser, C A , "Quota sampling," *Journal of the Royal Statistical Society*, A, 115 (1952), 411–23.

[44] Moser, C A , "The use of sampling in Great Britain," *Journal of the American Statistical Association*, 44 (1949), 231–59.

[45] Moser, C. A., and Stuart, A., "An experimental study of quota sampling," *Journal of the Royal Statistical Society*," A, 116 (1953), 349–405

[46] Moss, L , "Sample surveys and the administrative process," *UNESCO International Social Science Bulletin*, 5 (1953), 482–94.

[47] National Institute of Social and Economic Research, *Register of Research in the Social Sciences*," London: 1954.

[48] Palca, H., and Davies, I. G R , "Earnings and conditions of employment in agriculture," *Journal of the Royal Statistical Society*, A, 114 (1951), 50–8.

[49] Peaker, G F., "A sampling design used by the Ministry of Education," *Journal of the Royal Statistical Society*, 116 (1953), 140–59

[50] Silvey, R. E , "The methods by which the B.B.C. endeavours to measure viewing," *Transactions of the Manchester Statistical Society* (1950–51), 1–37.

[51] Stone, R., Utting, J. E. G., and Durbin, J., "The use of sampling methods in

national income statistics and social accounting," *Accounting Research*, 7 (1950), 333–56.

[52] Treasure, J. A. P., "The British Export Trade Research Organisation, 1945–52," *Manchester School of Economic and Social Studies*, 21 (1953), 118–39.

[53] Treasure, J. A. P., "Retail audit research," *The Incorporated Statistician*, (1953), 148–159.

[54] Utting, J. E G , Social accounts of local authorities Cambridge: Cambridge University Press, 1953.

[55] Utting, J. E. G., and Cole, D., "Sampling for social accounts—some aspects of the Cambridgeshire Survey," *Bulletin of the International Statistical Institute*, 34 (1954).

[56] Utting, J. E G., and Cole, D , "Sample surveys for the social accounts of the household sector," *Bulletin of the Oxford Institute of Statistics*," 15 (1953), 1–24.

[57] Wadsworth, R. N , "The experience of a user of a consumer panel," *Applied Statistics*, 1 (1952), 169–78

[58] Wilkins, L. T., "Criminology and Prediction," *Operational Research Quarterly*, 5 (September 1954), 80–6

[59] Wilkins, L. T , "Prediction of the demand for campaign stars and medals," *Central Office of Information*, London, 1949.

[60] Yates, F., *Sampling Methods For Censuses and Surveys*," 2nd edition, London. C. Griffin and Co. Ltd , 1953.

USES OF CRUDE VITAL RATES IN THE ANALYSIS OF REPRODUCTIVITY*

George J. Stolnitz
Princeton University

I. INTRODUCTION

Gross and net reproduction rates continue to be widely used in population studies, as they have been for the last quarter of a century. Although subjected to increasing criticism in recent years, such rates are still likely to be cited as uniquely valuable indicators of demographic behavior and prospects, while their unavailability for many areas is almost universally taken to be a major gap in knowedge. The purpose of this paper is to consider whether the kinds of information commonly sought from reproduction rates may be inferred from crude vital measures, which are much more generally obtainable. In particular, to what extent can it be expected that the two types of measures will show parallel variations over time or space?

In view of the importance usually attached to reproductivity patterns, the question is of obvious interest as a problem in estimation But in addition it throws new light on some methodological issues which have long been regarded as resolved Reproduction rates are supposed to provide a needed escape from crude rates and the possibility that the two might be substitutable in significant ways seems to challenge the very purpose of a large part of traditional demographic measurement. It is useful, therefore, to begin by reviewing the main arguments advanced in support of reproduction rates

Almost invariably, such arguments have started with the premise that crude vital measures present a distorted view of demographic behavior. A population's crude birth or death rate for any period depends not only on the rate of childbearing or dying in each age class, but also on the relative numbers of persons in such classes.[1] Since these numbers

---

* The author is indebted to Clyde V Kiser, George F Mair, Leighton van Nort and Norman B. Ryder for criticisms and to Erna A Härm, Blanche E Cogan, Elisabeth Souther and Hugh Dingle for computational assistance

[1] The birth rate for any calendar period is defined as the ratio B/P, where B is the average annual number of births in the period and P the average size of the population (generally expressed in thousands). Let b_i denote the births among women of age i, p_i the number of women at that age, r_i the age-specific fertility rate b_i/p_i, and f_i the fraction p_i/P. It follows that the birth rate can be written as

$$\frac{\Sigma\, b_i}{P} = \frac{\Sigma\, \frac{b_i}{p_i}\, p_i}{P} = \Sigma\, r_i f_i,$$

i e , as a weighted sum of the age-specific rates (The births could as readily be classified by age of father, but the criteria for preferring one classification over another are often complicated and need not concern us here.) The death rate can be represented in the same way, the weights at each age being the ratio of persons of both sexes to the total population

are the outcome of earlier movements in fertility, mortality and migration, any crude rate can describe the present only at the expense of also involving the past.

The avoidance of this difficulty has been cited as the great contribution of gross reproduction rates in the study of fertility. Any two populations with the same age-specific rates of childbearing will have identical gross rates, whatever the differences in their current age distributions. The corresponding property holds for the net reproduction rate, this time in connection with the analysis of replacement. The percentage rate at which a population will eventually change if existing vital conditions continue into the future can very generally be determined without any reference at all to its current age composition. The net reproduction rate provides a measure of this change, over time spans of a generation (approximately 30 years).[2]

It is typical of the large literature on these questions that it deals almost exclusively with the problem of analyzing the experience of a given time and place. Little or no theoretical attention has been given to considering the contrasts between crude and reproduction rates for purposes of studying trends or spatial variations. This oversight has been a natural one in some respects In the first place, reproduction rates give the appearance of being inherently dynamic, in that even a single value describes how the population may change over time Thus if a given experience indicates a net rate of 1.2, it also implies an eventual population growth of 20 per cent per generation, and this

---

[2] Exceptions to the last two statements are conceivable but are of mathematical interest only Actual populations subject to unchanging age-specific fertility and mortality rates will after a while take on a "stable age" form, whose age composition and rate of growth are constant over time. These are independent of the population's original age composition.

Not the least of the attractions of reproduction rates has been their apparent common-sense appeal Basically, the gross rate can be interpreted as measuring the number of children ever born to women who have already survived to the end of the reproductive period of life The net rate in addition to its function as a stable-age parameter, is intended to measure the number of children among women who are traced from the time of birth and may die before completing the childbearing years. Popular interest in reproductivity is likely to focus on questions requiring just these kinds of information. No doubt the natural tendency is to think of actual groups of women. By ready analogy, however, it is a small step to consider hypothetical groups, whose reproductive "history" summarizes the experience of women at different ages during the same calendar period. As before, let $\{r_i\}$ denote the set of age-specific fertility rates (per 1,000 women) observed in a given period. Then 1,000 women, assumed to survive from birth through the reproductive years and to have the observed fertility rates, would have r_i births at each age i. The total number of births during their lifetime would be Σr_i, the summation running from about age 15 to 50. The gross reproduction rate is defined as $S\Sigma r_i$ ($S\Sigma r_i \div 1{,}000$ if expressed on a per woman basis), where S is the ratio of daughters to total births (The reason for introducing this ratio is that it is often convenient to deal with parents and children of the same sex.) The net reproduction rate is defined as the number of daughters that would be born if the 1,000 women not only had the observed fertility rates but were also subject to the mortality risks of the period. Letting L_i denote the fraction who would survive from birth to age i under the given mortality conditions, the number of births at that age would be r_iL_i and the total children ever born Σr_iL_i. The net rate for the period would therefore be $S\Sigma r_iL_i$.

All of the reproduction measures used here stem from these definitions. No attempt has been made to consider the analogous measures occasionally published for actual population groups ("real cohort" rates), for males, or for groups classified by variables other than age (on the last see [12]).

figure can be compared with actual changes in the past. Secondly and more important, reproduction rates first came to be widely applied in the 1930's and to the special conditions of Western nations. In all of these areas long-run fertility trends had been very predominantly downward before World War II. On the then plausible assumption that these trends would continue and would more than offset further declines in mortality, the net reproduction rate for any current period could be regarded as an upper estimate of the rates of growth that would actually take place in the future. In particular, the fact that most Western net rates of the time were close to or below unity seemed clearly to presage eventual population decline throughout the region. To many this seemed much more important than the observation that the net rates themselves varied from period to period. The continued excess of birth rates over death rates in most instances only added emphasis to this line of thinking. A single net rate often appeared to tell more about the future than a whole time series of crude vital measures.

As a practical matter, however, reproduction rates have been used mainly in comparative contexts, the usual interest being less in individual values than in the way these measures have varied over time or space. In this sense, the theory just summarized has been largely irrelevant to application. Granting that net or gross rates have unusual advantages for describing a single experience, it is by no means clear that similar advantages hold for comparative purposes. This is not, of course, to deny the usefulness of documenting variations as accurately as possible. Indeed, the importance of such variations is taken for granted for present purposes,[3] Otherwise there would be little point to seeing if adequate approximations may be obtained by indirect means. The primary purpose of this paper is to examine the probable extent of similarity between movements in gross and net rates on the one hand, and crude vital rates on the other. Nevertheless, many of the following findings are also implicit criticisms of the traditional theory, not so much for what it says as for what it overlooks.

II. TEMPORAL VARIATIONS

Very nearly all of the reproduction rates discussed below were taken from publications by Depoid, Kuczynski and the United Nations [1, 6, 10]. The time series derivable from these sources cover 27 countries and come close to exhausting the existing materials for national

[3] Whether this is justified is another matter. The fact is that reproduction rates are still primary tools in practice. The critical literature, although extensive, seems to have ignored the question raised here, i.e., whether the practical uses of reproduction rates have really marked a significant departure from older and simpler measures. (See for example [2, 5, 9, 13].)

populations through the late 1940's, the terminal years for the data of this study.[4]

Crude birth and death rates were obtained for each of the periods for which a reproduction rate was available for an area and the matched values used to compare changes between periods. For example, the reported gross reproduction rate for France was 1.31 in 1904–07 and 1.23 in 1908–13, and the corresponding crude birth rates 20.4 and 19.3; in this case a decline of 6.1 per cent in the one measure could be compared with a decline of 5.4 per cent in the other. In view of the dimensional dissimilarity between the two types of measures, all variations have been expressed in percentage terms.

1. The gross reproduction rate. Since this rate is intended to describe fertility proper, its obvious analogue among crude vital measures is the birth rate. The matching process just described yielded a total of 460 paired changes between the successive periods on record in the various areas. Less than 6 per cent of such changes showed the measures moving in opposite directions, the value of 7.4 per cent in the first column of Table 1 was obtained by adding half the number of instances in which a non-zero movement in either measure was paired with a zero variation in the other. Small as they are, these fractions greatly exaggerate the significance of the apparent dissimilarities. As can be seen from Figure 1 (second and fourth quadrants), both measures varied by less than one per cent in over a third of the cases involving divergent movements. Almost always, therefore, a variation in the birth rate would have been sufficient to infer correctly that the gross rate moved in the same direction or that it failed to change appreciably.[5]

With respect to magnitudes, the similarity in the two sets of variations is brought out by the marked concentration of points about the main diagonal of Figure 1 and again by the position measures shown in Table 1. More than half of the paired changes differ by less than one percentage point and nearly four fifths by less than two points. Also,

---

[4] Many of the rates are for single years, but a large proportion are for longer periods. Partly because of this and partly because the time series are often incomplete the time intervals between successive measures tend to be highly variable within a country as well as between countries. The reliability of the rates and the ways in which they were computed also tend to vary widely. Some of the nineteenth-century data, taken from Depoid, were obtained by methods which exaggerate the apparent parallelism of movements in crude and reproduction rates. There is little reason to suppose, however, that these factors have had any substantial overall effects. If the data are deficient in any serious way, it is with respect to the underrepresentation of the populations of Latin America, Africa and Asia.

[5] Eleven of the countries in the sample had 20 or more changes and accounted for nearly three fourths of the paired variations. Only three of these areas had appreciable proportions of divergent movements, some 13–14 per cent, while seven had proportions below 6 per cent.

TABLE 1

SUMMARY OF DISTRIBUTIONS OF ABSOLUTE DIFFERENCES BETWEEN PAIRED PERCENTAGE CHANGES IN REPRODUCTION RATES AND CRUDE VITAL RATES*

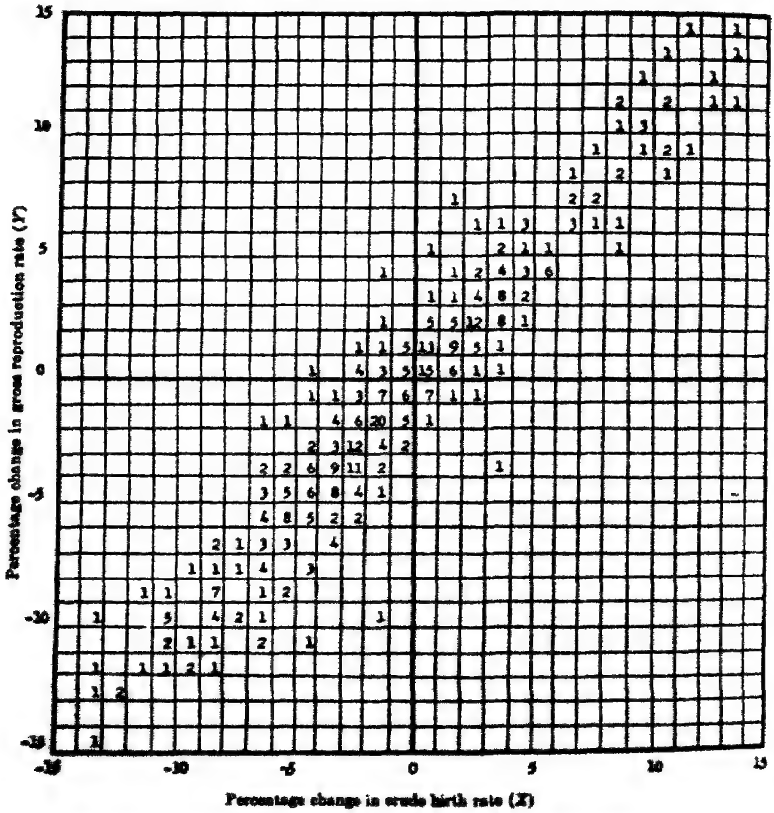
Item	GRR and BR	NRR and BR	NRR and combination of BR and DR†
1. All changes between successive measures			
No. of paired changes	460	431	431
Divergent movements (%)	7.4	12.5	24.6
Absolute differences between paired changes:			
Mean	1.3	2.3	5.3
First quartile	4	5	1.8
Median	.8	1.2	3.8
Third quartile	1.7	2.6	6.7
2. Pre-1900 changes between successive measures			
No. of paired changes	75	65	65
Divergent movements (%)	12.7	23.8	15.4
Absolute differences between paired changes:			
Mean	1.2	3.5	3.4
First quartile	.3	.8	1.6
Median	7	2.8	2.7
Third quartile	1.7	5.5	4.3
3. All 30-year changes			
No. of paired changes	112	88	88
Divergent movements (%)	8.9	20.5	21.0
Absolute differences between paired changes:			
Mean	4.4	10.2	9.1
First quartile	1.7	5.8	3.6
Median	3.8	9.2	7.8
Third quartile	6.6	13.1	14.4

* BR = Crude birth rate, DR = Crude death rate, GRR = Gross reproduction rate, NRR = Net reproduction rate

† The first of the combinations defined in footnote 18.

the differences in the total sample of changes have very nearly the same mean and quartile values as do the differences in the pre-1900 subsample. It follows that the distributions of differences before and after 1900 were practically indistinguishable in these respects. Only the proportions of divergent movements differed notably between the two periods, being almost twice as large before 1900 as subsequently. The greater relative frequency of such movements in the earlier data might have been expected to result in larger average differences as well. That this did not happen is explained by the tendency for both meas-

FIG. 1. Tabulation of matched percentage changes in the gross reproduction rate and crude birth rate.*



* Using changes between all successive periods on record in 27 countries. Changes of over 15 per cent in either measure occurred in about 1 of 10 cases. Estimating equation (all changes, ungrouped data): $Y = -.12 + 1.04X$.

ures to change very little when varying in opposite directions. In each period the differences encountered in such instances had the effect of lowering, not raising, the average.

These patterns are especially interesting in that the pre-1900 variations usually spanned intervals of 5 or 10 years, while those for later periods were mainly annual changes. Thus it appears that the birth rate is almost as useful an estimator for intermediate periods as for very short-run intervals. This seems to be true even if we take account of the likelihood that some of the pre-1900 differences between the two measures are downward-biased (see footnote 4).

To test for the kinds of relationships to be expected for longer periods, use was made of all the 30-year changes that could be derived from the original time series For example, if gross reproduction rates in an area were available for 1901–05 and 1931–35, and again for 1906–10 and 1936–40, both changes were computed and paired with the concurrent movements in the birth rate. The results are summarized in the bottom tier of Table 1.

The larger errors generally encountered are in accord with expectations. Differences between the percentage movements of the two measures are likely to be substantially greater over periods of a generation than over intervals of a year or even a decade.[6] Nevertheless, the present results seem to the writer further evidence that the movements are sufficiently parallel for most practical purposes. In the first place, changes between dates as far apart as three decades are uncommon in actual studies of fertility trends. The reasonable presumption is that the average errors for such frequently cited intervals as 1 to 2 decades would be appreciably lower. Secondly, the errors to be expected should be considered in relation to the length of the period in question. The comparatively short-run changes in the gross reproduction rate which were discussed earlier had a mean absolute value of a little over 6 per cent, whereas the 30-year changes averaged over 21 per cent In relative terms, therefore, an average 30-year error of about 4 percentage points is not much different from the previous results. Errors of this size would be prohibitive for purposes of estimating short-period movements, for example within a decade, but could hardly distort perspectives seriously in dealing with trends spanning a generation.[7]

[6] The longer the interval of change, the greater are likely to be the differences between the age-composition factors affecting the change in the birth rate. In contrast, constant "population weights" of unity are used in deriving the gross reproduction rate (see footnote 2 and also Tables 3 and 4 below).

[7] Thus if the gross reproduction rate declined by 30 per cent in country A and by 33 per cent in B, it is quite possible that the trends of the birth rates would make the change appear to be more rapid in A. However, differences as small as 3 or even 5 percentage points over periods as long as 30 years are not likely to be singled out for special attention in comparative studies. In fact, such differences may

Thirdly, long-run movements of the two measures are very likely to have the same direction. No instance of a divergent 30-year movement was encountered for any period beginning after 1860 and the concentration of such movements in earlier periods is probably attributable in good part to inaccurate data. So far as the present materials show, a practically unerring rule of thumb would be that a 30-year movement in the birth rate which exceeds 5 per cent implies a similar direction of the 30-year trend in the gross rate. A change below 5 per cent in the birth rate can be taken to imply that, although the movement in the gross rate is of uncertain direction, its magnitude is unusually small.[8]

Obviously such arguments could be pushed too far, even if we could assume that the present data are more comprehensive than is in fact the case. Certainly it would be well to allow for the likelihood that use of the birth rate as an estimator leads to rising average errors, the longer the interval of change. Moreover, whether the errors indicated here are prohibitive or tolerable depends on the objectives of estimation and these are rarely the same in different studies. Nevertheless, it does seem decisive that the suggested magnitudes are no larger than the errors frequently implicit in direct measures of the trend in the gross rate. It would be difficult to cite instances in the literature in which such magnitudes have been the occasion for any significant conclusions.

This interpretation, if justified, has several important implications: (1) There are many contemporary populations for which birth rates are available but whose trends in the gross reproduction rate cannot be ascertained directly. Reasonable approximations are the most to be expected in these instances and general orders of magnitude are often enough. It can be expected that the main features of the unknown trend between any desired periods can be inferred from the direction and magnitude of the corresponding percentage changes in the birth rate. The latter changes may also serve in a second capacity, as useful checks on the reliability of gross rates estimated by other indirect methods.[9] (2) Use of birth-rate changes as estimators or checks should prove similarly helpful in numerous historical investigations. (3) In most statistically advanced areas, reproduction rates are published a few years after the event, whereas birth rates are generally known almost immediately. In such situations the birth rate may serve not only to estimate the change in the gross rate but also its most recent value.

---

will be grouped under the same class interval in compiling distributions of international movements. The more important aspect, that A and B both experienced unusually rapid declines, would almost certainly be brought out by their birth rates.

[8] In all but one divergent movement, the birth rate varied by less than 5.5 per cent. Conversely, all such movements involved changes of 5 per cent or less in the gross rate and all but two involved changes below 3.5 per cent.

[9] See for example [4] and [8].

These applications refer mainly to the estimation of single or occasional changes. There is impressive evidence that the birth rate can also be used to reconstruct the salient aspects of entire time series of gross rates. Thus among the eleven countries with 20 or more values on record, the coefficient of linear correlation between the two measures was .93 in one instance, 98 to 99 in four cases, and .99 or over in the remaining six. The implications of these results for purposes of estimation are obvious. Perhaps equally interesting is their bearing on traditional methodology. Judging from the past, at least, our substantive knowledge of movements in the gross reproduction rate would have been very nearly the same, had it been necessary to rely on the birth rate alone. It is slight exaggeration to say that the theory underlying the gross rate has been less important as a source of new empirical findings than as a basis for reinterpreting the trend of the birth rate.[10]

2. The net reproduction rate Since changes in this measure reflect variations in both fertility and mortality, the crude rate of natural increase (birth rate minus death rate) suggests itself most immediately as the appropriate counterpart. The evidence is convincing, however, that this is much too simple a view. The present data permitted a pairing of some 430 changes between successive values of the two rates; most of the variations were over short intervals, averaging under 5 years. With respect to direction, the changes are quite similar Divergent movements were encountered in only 1 of 5 instances and often involved comparatively small changes on the part of both measures. Moreover such movements would have been much less frequent, had the published net rates been more accurate. In many cases these were computed on the basis of non-current mortality measures (the L_t of footnote 2); as a result, the published rate was often declining when the true rate was in fact rising along with the crude rate. Examination of the paired variations over 30-year periods, which were much less affected by these distortions, showed a considerably lower proportion of divergencies, 1 in 8

The crude rate of increase seems even more appropriate for judging whether the net rate is above or below unity As observed earlier,

[10] The size of the above correlations and the small differences between the paired short-run changes discussed previously are by no means unrelated. This can be suggested most easily by assuming the birth and reproduction rates are differentiable functions of time, say $x(t)$ and $y(t)$, respectively. Since $y=0$ when $x=0$ (i.e., when no births occur), perfect correlation between the two measures would imply

$$y = cx, \quad \frac{dy}{dx} = c = \frac{y}{x}, \quad \text{and} \quad \frac{dy}{y\,dt} = \frac{dx}{x\,dt},$$

or zero differences between the paired percentage variations over time. Conversely, the last condition implies perfect correlation, with the regression line passing through the origin. In approximate terms, therefore, very high correlations between the two measures are necessary and sufficient conditions for the expectation that their short-run percentage changes will be closely similar.

this question has attracted especially widespread interest in the past. Of the approximately 160 net rates below unity in the present compilation, 60 per cent were paired with a crude rate whose absolute value was below 5 per 1,000, and 94 per cent with one below 8 per 1,000. And conversely, 5 out of 6 crude rates below 5 per 1,000 and 3 of 4 below 8 per 1,000 were paired with fractional net rates.

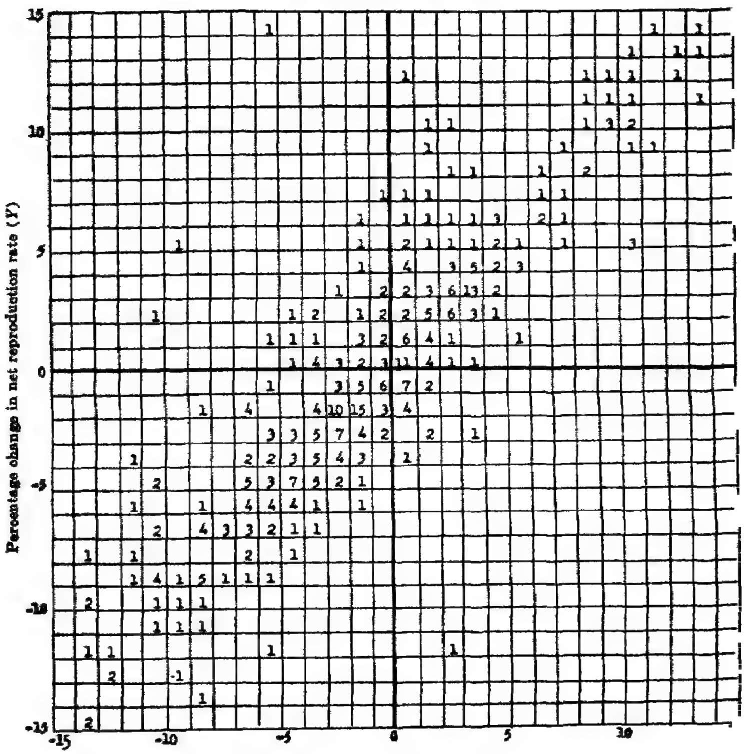
On the other hand, crude rates of increase are likely to be of little or no value for estimating magnitudes of change. The distribution of absolute differences between the 430 paired changes showed a mean of 51 percentage points and a median of nearly 9 points. The major explanation of the huge mean difference was that many of the crude rates were close to zero. In such instances even small absolute changes could represent tremendous relative movements, often amounting to several hundred per cent. Thus the mean relative change in the crude rate was over 55 per cent, in contrast to the average of only 6 per cent for the much more stable net rate. After eliminating all crude rates under 5 per 1,000, it was found that the remaining paired changes had a mean difference of about 10 percentage points; the median also dropped appreciably, to somewhat less than 6 points. As a practical matter, however, it is doubtful that even this restricted use of the crude rate has much value, since the differences to be expected appear far too high relative to average changes in the net rate. And in any event, low rates of natural increase, with a resulting instability in their percentage variations, will continue to occur frequently in the future.

Accordingly, several attempts were made to pair the changes in the net rate with more stable estimators. As a first approximation, use was made of the birth rate, although it could be anticipated that the estimates would be biased. Mortality trends being very persistently downward among the populations in the sample, the net rates tend to rise more or fall less than if affected by fertility alone.[11] As a result, the percentage variations in the birth rate tend to be algebraically lower than those in the net rate. This is brought out clearly by Figure 2. Among the paired changes moving in the same direction (first and third quadrants), the movement in the birth rate was more negative or less positive in the large majority of instances. Similarly, two thirds of the divergent movements involved a rise in the net rate and a fall in the birth rate.

[11] Any decline in mortality before the end of the childbearing period of life has the effect of raising the net rate; any rise lowers it. In contrast, the effects of mortality trends on the birth rate tend to be both indirect and delayed, since they operate mainly through changes in the sex-age distribution of the population.

Nevertheless, the association between the two sets of variations is striking on the whole, as can also be seen from the middle column in Table 1. Nearly half of the paired changes differed by less than 1 percentage point and two thirds by less than 2 points. Errors of these magnitudes are well within the tolerance limits required for most purposes of estimation In specific problems, moreover, the likely direction of error can often be taken into account. Thus if the birth rate is rising and the death rate falling, it will generally be safe to assume that the

FIG. 2.—Tabulation of matched percentage changes in the net reproduction rate and crude birth rate.*



* Using changes between all successive periods on record in 27 countries. Changes of over 15 per cent in either measure occurred in about 1 of 10 cases. Estimating equation (all changes, ungrouped data): $Y = 1.33 + 1.05\,X$

percentage change in the former measure is below the change in the net rate. The frequency of divergent movements was about 1 in 8 cases and the proportion falls to 1 in 10 after eliminating the instances in which one of the paired measures showed no change at all. In addition, many of the divergencies involve unusually small variations in both measures (see Figure 2).

These conclusions should be interpreted in the light of the fact that most of the changes refer to short-term intervals, between 1 and 5 years Table 1 makes it clear that the usefulness of the birth rate as an estimator declines rather rapidly as the intervals lengthen. The above errors for the total sample are appreciably smaller than the ones shown for the subsample of changes occurring before 1900, which generally cover 5- or 10-year periods. In turn, both of these sets of errors are well below the differences that seem indicated for 30-year periods. These patterns are not accidental, the basic explanation being the role of mortality. The longer the interval of change, the greater is likely to be the extent to which the net reproduction rate has been affected by mortality movements. The birth rate being comparatively insensitive to such trends, the two measures can be expected to show increasingly dissimilar variations with the passage of time.[12]

Somewhat surprisingly, no simple combination of birth and death-rate changes could be found which would overcome these difficulties. Three combinations were investigated, the one which in general yielded the smallest errors being described in the last column of Table 1.[13] It will be observed that the birth rate alone seems a much more serviceable estimator for short-run purposes. Neither the birth rate alone nor any simple combination with the death rate stands out as clearly superior in the pre-1900 changes, a result which may be attributable to the small number of cases.[14] Only for the 30-year changes is there any strong suggestion that simple methods of allowing for mortality may

[12] So far as the populations of the West are concerned, it can also be expected that the future effects of mortality trends will be much smaller than in the past. Current mortality before age 45 is so low in these areas that the maximum effects still possible from further declines are limited.

[13] Letting B and D denote the percentage changes of the birth and death rates, respectively, the three combinations were

$$\frac{(100+B)(100-D)] - 10{,}000}{100}, \quad \frac{100+B}{100+D} \times 100 - 100, \quad \text{and} \quad B - D.$$

The first of these is the one whose pairings with the net rate are described in Table 1. It will be noted that each combination reduces to B when $D=0$, also, each yields a higher value than B when D is negative.

[14] Examination of all the paired 5-year changes that could be derived from the available materials yielded these results: an average change in the net rate of 13 per cent, a median error of 2.7 per cent resulting from the use of the birth rate alone, and a median error of 5.2 per cent from the combination discussed in the text. The corresponding values for all possible 10-year changes were 17, 4.7, and 5.3.

lead to sensibly smaller errors, and even here there are contrary indications. Considering that the average 30-year change in the net rate was 15–16 per cent, the errors to be expected from all of the estimators examined here are probably too high for most purposes.

At the same time, it should be added that the present evidence is of a very limited sort. Much smaller errors might be found by combining the changes in birth and death rates in a less simple manner, for example by differential weighting The important analytical lead in this regard may be that, although the net rate tends to vary in fairly proportionate manner to a given change in the birth rate, its relation to the death rate is much more complicated Pending actual investigations, however, it is impossible to predict the kinds of errors that might be expected from a more sophisticated approach.

Much more can be said about the potential usefulness of crude vital measures in a second connection, i.e , for inferring the over-all characteristics of time series of the net rate. Among the eight countries for which 20 or more values were available, the lowest coefficients of linear correlation between this measure and the crude rate of natural increase were .83, .87, and .91; in the remaining five areas, the coefficients were all between 95 and .98 The birth rate for the same countries yielded coefficients of .66, .88, .93, and in five instances, .95 to 99. The low value of 66 is for France, where a succession of wars over the last century has led to unusual imbalances in the sex-age distribution of the population In general, therefore, the conclusions reached earlier about the gross reproduction rate also seem appropriate here, though to a lesser extent As a contribution to time series analysis of population replacement, the theory underlying the use of net rates has been primarily fruitful along conceptual lines. Given the theory, much the same factual interpretations would have been forthcoming from other empirical sources.

In summary, the results of this section suggest the following: (1) The crude rate of natural increase is probably of little practical value for estimating single or occasional percentage changes in the net reproduction rate (2) On the other hand, extended time series of the former measure provide reliable information on at least the salient time-series characteristics of the latter. (3) The crude rate of increase also permits confident inferences on whether the net rate is close to or below unity and whether the net rate is rising or falling (4) For estimating short-run percentage variations, not exceeding a decade, it is likely that changes in the birth and net rates are closely enough associated to be substitutable for most practical purposes. Perhaps the main

reservation here is that many of the world's statistically underdeveloped populations may experience very rapid mortality declines during the next few decades; the postwar declines found in numerous parts of Latin America and Asia are almost without precedent in the earlier vital history of Western nations. (5) No simple combination of birth and death rates is likely to yield better short-run estimates than the birth rate alone. (6) None of the estimators examined here seems satisfactory for dealing with changes in the net rate over intervals exceeding a decade, unless very crude indications are sufficient. On the other hand, further investigation may well reveal combinations of birth and death rates which are more serviceable.

Obviously, the prospects that crude vital measures can provide useful inferences about the net reproduction rate are mixed. There is little question, however, that such inferences will often be possible in the kinds of situations discussed earlier in connection with the gross rate.

III SPATIAL VARIATIONS

The usefulness of crude vital measures for inferring contemporary international differences in gross and net rates has been examined for three periods, centering on 1933, 1937, and 1947. Ideally, the question should have been reviewed for a large number of periods and for significant sub-groupings of areas. The paucity of data was such, however, that only recent years could be used at all; even then the number of areas for which documentation exists at any one time is small and non-Western populations are badly underrepresented. In each year correlation coefficients were used as indexes of the association between crude and reproduction measures.

Two aspects of the testing procedure should be noted. First, the areas selected for 1933 and 1937 are practically identical. This was done intentionally, a few countries being rejected which were documented for one year but not the other. Thus the results may be interpreted as the outcome of a crude replication experiment; the relatively local changes in each area between the two years correspond to sample-to-sample fluctuations and the inclusion of the same areas corresponds to repeated sampling from the same universe. Secondly, all of the countries for which the needed data were available were included for 1947. About a third of the 1947 areas are not in the earlier samples while about a quarter of the areas in the early years do not appear in the later year. In view of these variations and the substantial differences between economic and demographic conditions before and after the

war, the results presented in Table 2 reflect about as much year-to-year variability as can be obtained.

As with temporal variations, the spatial associations between birth and gross reproduction rates appear to be very close. Whether measured in rank order or cardinal terms, the correlations suggested by the present data are well over .9 and highly similar from period to period. If the inadequacies in coverage are not too serious, it seems clear that the birth rate can be confidently applied to two kinds of problems: (1) the reconstruction of entire comparative patterns, when direct information on the gross rate is lacking for most of a group of countries under study, and (2) the determination of the comparative status of individual populations, in relation to areas which provide such information.

These conclusions are brought out in a more analytical way by Table 3, which presents selected data on sex-age composition and age-specific fertility (corresponding to the f_x and r_x of footnote 1). The five populations for which recent fertility rates are given span a broad range of demographic situations, while the earlier materials for Sweden and the Netherlands suggest the compositional changes that may take

TABLE 2

CORRELATION OVER SPACE BETWEEN REPRODUCTION RATES AND CRUDE VITAL RATES*

Correlates	No of countries	Coefficient of rank order correlation	Coefficient of linear correlation
GRR and BR			
1933	18	.93	.95
1937	18	93	98
1947	20	94	97
NRR and CRNI			
1933	17	.89	.88
1937	18	.83	.73
1947	20	.83	88
NRR and BR			
1933	17	.82	.78
1937	18	.84	.76
1947	20	.74	.76

* BR = Crude birth rate, CRNI = Crude rate of natural increase, GRR = Gross reproduction rate, NRR = Net reproduction rate.

place over periods of very marked vital trends. It will be observed that the variations in sex-age proportions are considerably smaller than those in fertility. Thus the latter measures, which completely determine the gross reproduction rate, also dominate the birth rate.

The expectation that the two rates will show similar patterns over time or space follows from these relations. Table 4 shows the birth rates resulting from all possible combinations of the sets of age-specific fertility rates and sex-age proportions in Table 3. The differences within each column, reflecting the influence of variable age composition, are usually far below the differences between the actual birth rates. There is no overlap at all between the ranges of values in the first four columns and five of the seven values in column 5 exceed the highest in column 4. In other words, a random selection of five values from the various columns would show the same or practically the same ordering of birth rates as was in fact the case. In turn, the ordering of gross rates indicated by such a selection would almost certainly correspond to the one suggested by the actual birth rates, which is correct.

TABLE 3

RATIOS OF NUMBERS OF FEMALES IN REPRODUCTIVE AGES TO TOTAL POPULATION AND AGE-SPECIFIC FERTILITY RATES, FOR SELECTED COUNTRIES AND PERIODS

Country, year	Age						
	15–19	20–24	25–29	30–34	35–39	40–44	45–49
	1. Ratio of females to total population (in per cent)						
Sweden							
1900	4.62	4.09	3.50	3.07	3.20	2.99	2.59
1950	2.92	3.25	3.75	3.81	3.82	3.81	3.50
Netherlands							
1909	4.72	4.28	3.86	3.61	3.14	2.78	2.42
1949	3.98	3.93	4.02	3.40	3.46	3.30	3.00
Japan, 1950	5.10	4.69	4.03	3.41	3.22	2.74	2.38
Venezuela, 1950	5.01	4.72	3.94	3.13	2.88	2.28	1.73
El Salvador, 1950	5.48	5.03	3.98	3.09	3.10	2.43	1.88
	2. Age-specific fertility rates (per 1,000 women)						
Sweden, 1950	38.9	126.0	128.0	93.6	54.9	18.3	1.5
Netherlands, 1949	12.7	97.0	187.8	172.0	120.2	49.5	4.4
Japan, 1950	13.3	160.3	236.9	174.8	104.2	36.0	2.3
Venezuela, 1950	104.4	276.4	277.5	217.3	144.7	52.4	20.1
El Salvador, 1950	120.0	296.0	320.2	249.4	146.0	58.4	18.3

TABLE 4

BIRTH RATES UNDER VARYING COMBINATIONS OF SEX-AGE PROPORTIONS AND AGE-SPECIFIC FERTILITY RATES, FOR SELECTED COUNTRIES AND PERIODS*

Sex-age proportions of:	Age-specific fertility rates of:				
	Sweden, 1950	Netherlands, 1949	Japan, 1950	Venezuela, 1950	El Salvador, 1950
Sweden					
1900	16.6	21.8	25.3	39.2	43.4
1950	16.4†	23.8	26.6	38.9	43.1
Netherlands					
1909	17.8	23.5	27.3	41.8	46.3
1949	17.4	23.6†	27.2	40.9	45.3
Japan, 1950	18.5	24.0	28.1†	43.5	48.1
Venezuela, 1950	17.9	22.7	26.9	41.7†	46.3
El Salvador, 1950	18.6	23.4	27.8	43.5	48.2†
Gross reproduction rate	1.10	1.56	1.76	2.66‡	2.95‡

* Each figure is the sum of the products obtained by cross-multiplying the indicated row in the first part of Table 3 by the indicated row in the second part.

† Differs somewhat from actual birth rate

‡ Estimated, but must be very nearly accurate, no official figure available

Once again, the advisability of relying on similar information in the case of the net rate is more uncertain. The associations of this measure with either the crude rate of natural increase or the birth rate are notably lower, while the year-to-year variations tend to be larger. An added complication is the questionable quality of some of the data; the numbers of cases are so small that even a few inaccuracies may have substantial effects on the indicated results. Fortunately the errors resulting from this factor appear to have operated mainly to reduce the apparent association.[15] On the other hand, the populations whose data are most suspect are also the ones which represent demographic situations outside the West. Thus, whether crude rates can be usefully applied to comparing net rates in the West and elsewhere must remain a moot point in good part.

Nevertheless, two positive conclusions seem rather clearly indicated.

[15] With only one country, Chile, removed from the 1947 data, the rank order coefficient between net and birth rates rises from .74 to .87 and the linear coefficient from .76 to .92. Had the net rate for that country been based on current mortality, the 1947 coefficients in Table 2 would have been intermediate between these values.

On the average and without distinction among sub-groupings of populations, the use of crude rates for judging international variations in the net reproduction rate will generally rest on correlation coefficients of about .8 or above. These are very high levels of association by usual standards in social studies. Secondly, the coefficients between birth and net rates in Table 2 probably understate the associations that can be counted on in dealing with Western populations alone. Mortality differentials within this region are already small and seem certain to become smaller. For this reason alone, it appears likely that coefficients below .8 will be unusual in the future. In addition, current mortality before age 45 in nearly all of the West is so low that future declines will leave only slight room for differences between gross and net rates. In the years to come, therefore, correlations between birth and net rates should increasingly approach the much higher associations in evidence today between birth and gross rates.

IV. CONCLUSIONS

In a number of ways, empirical use of gross and net reproduction has represented something of a shotgun methodology: (1) Reproduction-rate theory has really been concerned with the experience of a given time and place. As applied, however, the theory has been assumed to demonstrate that the disparities between crude and reproduction rates are equally significant for the very different purposes of making spatial or temporal comparisons. (2) Most of the theory has in fact centered about the advantages of the net reproduction rate. The tendency in empirical work, on the other hand, has been to proceed as if the break with crude vital measures is similarly marked in the case of the gross rate. Whether this is so, and what the dimensions of a significant break might be, are questions which have been answered more by default than by design. (3) The frequency with which reproduction rates have been computed to the third decimal place would make it appear that they are to be used in very precise ways. In contrast, the examples most stressed by the theory have been of an ordinal kind, or situations in which the net rate appeared to point to imminent population decline and the crude rate of natural increase to continuing growth.

The results presented earlier bring out the relevance of these distinctions. There can be little question that the development of reproduction-rate theory has fostered a richer conceptual interpretation of demographic events than was previously possible. But it also seems clear that the specific factual contents of such interpretations have

often been derivable, with sufficient precision, from much simpler vital measures. In particular, undue emphasis seems to have been placed on the contrast between positive current replacement, as indicated by current natural increase, and negative eventual replacement, as suggested by fractional values of the net reproduction rate. Below-replacement values of the net rate have far more often than not been accompanied by current rates of increase which are unusually low by historical standards. And conversely, the large majority of such current rates could have been used to infer that the net rate was below unity. The net rate measures eventual population change only on the assumption that age-specific fertility and mortality remain the same over lengthy periods, usually two generations at least Given the same assumption, the onset of negative crude rates of increase could be deduced from much shorter projections of current behavior into the future.

Issues in methodology as broad as the ones raised here can never, of course, be resolved definitely Two reservations in particular need emphasis in interpreting the previous conclusions First, any judgment as to what may or may not constitute significant errors is inevitably subjective in some degree. The criteria used above seem pertinent to the bulk of demographic measurement but others could plausibly be advanced, at least for some purposes The second and more important point is the underrepresentation of non-Western materials in the present evidence. The nineteenth-century experience on record serves as a partial substitute for contemporary non-Western situations, this is one of the reasons why it has been singled out for special attention. Nevertheless, this experience relates predominantly to Western Europe and it is advisable to exercise caution in applying the above findings to the statistically underdeveloped areas of today.

REFERENCES

[1] Depoid, Pierre, *Réproduction nette en Europe depuis l'origine des statistiques de l'état civil.* Published under France, Statistique Générale, Études Démographiques, No. 1 (Paris: Imprimerie Nationale, 1941), 39.

[2] Dorn, Harold F., "Pitfalls in population forecasts and projections," *Journal of the American Statistical Association*, 45, No. 251 (September 1950), 311–34.

[3] Dublin, Louis I. and Lotka, Alfred J., "On the true rate of natural increase of a population," *Journal of the American Statistical Association*, 20, No. 151 (September 1925), 305–39.

[4] Glass, D. V., *Population Policies and Movements.* Oxford: Clarendon Press, 1940, Appendix.

[5] Hajnal, John, "The analysis of birth statistics in the light of the interna-

tional recovery of the birth-rate," *Population Studies*, 1, No. 2 (September 1947), 137–64.

[6] Kuczynski, Robert R., *The Measurement of Population Growth*. New York: Oxford University Press, 1936, Tables 37, 57, and 58.

[7] Lotka, Alfred J., *Théorie analytique des associations biologiques*, Part II. No. 780 in the series Actualités Scientifiques et Industrielles (Paris: Hermann & Cie., 1939), especially 64–77.

[8] Shaul, J. R. H., "Derivation of total fertility, gross and net reproduction rates from census statistics of marriage fertility," *Journal of the Royal Statistical Society*, 109, Part 3 (1946), 278–83.

[9] Stolnitz, George J. and Ryder, Norman B., "Recent discussion of the net reproduction rate," *Population Index*, 15, No. 2 (April 1949), 114–28.

[10] United Nations, *Demographic Yearbook, 1949–50*, Table 24.

[11] Whelpton, P. K., *Cohort Fertility*. Princeton: Princeton University Press, 1954, Ch. 1.

[12] Whelpton, P. K., "Reproduction rates adjusted for age, parity, fecundity, and marriage," *Journal of the American Statistical Association*, 41, No. 236 (December 1946), 501–16.

[13] Wolfenden, Hugh H., *Population Statistics and Their Compilation* (Rev. ed.). Chicago: University of Chicago Press, 1954, Ch 14.

DIFFERENTIAL MORTALITY, GENERAL AND CAUSE-SPECIFIC IN BUFFALO, 1939–41

CONSTANTINE A YERACARIS
The University of Buffalo

INTRODUCTION

DEMOGRAPHIC phenomena provide the social scientist with relatively precise and objective data. Research shows such phenomena closely related to social and cultural factors. Age, nativity, race, sex and socio-economic status relate to mortality [2, 5] the relation varying with the cause of death [1, 8, 9].

This paper presents further evidence of variation of demographic phenomena with socio-economic status. It summarizes highlights of a study [12] undertaken to combine and retest previous methodological and substantive findings concerning mortality in a metropolis. It presents for the city of Buffalo, New York: (1) differentials in mortality from all causes; (2) cause-specific mortality differentials, and (3) infant mortality differentials.

PROCEDURES AND METHODS

Data on deaths in 1939–41 of Buffalo residents were secured from records of the Division of Vital Statistics of Buffalo and from records of the New York State Health Department. Death rates were computed by averaging the number of all age-specific deaths of Buffalo residents for the three-year period and relating them to the age, sex, race and nativity composition of the population of Buffalo as reported in the 1940 Population Census. Infant mortality rates were based on the three-year average number of deaths of those under one year of age applied to the average number of resident births for the same period, adjusted to the year 1938

Crude, standardized, and age-specific death rates and life expectancies have been basic tools of this mortality analysis [12]. In addition, for a simpler summarizing measure of socio-economic differentials in mortality, the simple, general, and proportional-cause-specific indexes of differential mortality were used These indexes are a modification of Westergaard's "method of expected cases" [11] The *simple index* measures the proportion of total deaths (from all or specific causes) within a given economic group in "excess" because of differences in mortality rates between this group and the highest economic group, with age held constant between these groups. It is the ratio of the difference between observed and expected deaths in this given group

(with age-specific mortality rates of the highest group applied to the age composition of this group) and the total observed deaths. For example, the simple index of Group II is given by

$$1 - \frac{\sum_{a=0}^{\infty} m_a{}^{I} n_a{}^{II}}{\sum_{a=0}^{\infty} m_a{}^{II} n_a{}^{II}}$$

where the superscripts denote socio-economic groups, the subscripts denote age groups, m denotes mortality rates, and n denotes the number in an age group.

The *general index* measures the proportion of all deaths (from all or specific causes) of the population as a whole which would have been "saved" if the age-specific mortality rates of the group with the lowest mortality prevailed in all groups.

The *proportional cause-specific index* measures the proportion of deaths from all causes that would have been "saved" if there were no economic differences in mortality from a specific cause. It is the ratio of the difference between the observed and the expected deaths from a given cause (with the death rates from the given cause within the highest economic group prevailing throughout the age composition of all economic groups) to the total observed deaths from all causes in all groups.

The five socio-economic groups for which rates were presented were obtained by classifying the 72 census tracts in the city according to the median (contract or estimated) rental of residences in the tracts The method was originally used by Hauser [3] and later utilized by Convis [1], Mayer [5] and Kitagawa [4] in their studies on mortality and fertility differentials for Chicago. The original rental intervals for Chicago (1930) were: Group I (highest), $75.00 and over; II, $60.00 to $74.00; III, $45.00 to $59.00; IV, $30.00 to $44.00, and V (lowest), under $30.00.

Since these rental intervals were not applicable to Buffalo for 1940, the 1940 population of Buffalo was divided into five groups in as nearly as possible the same percentage distribution as the Chicago intervals. The 1940 census figures for contract or estimated median rentals provided the following socio-economic intervals for Buffalo. Group I (highest) $42.50 and over; II, $31.90 to $42.00; III, $24.70 to $31.89; $20.00 to $24.69; and V (lowest), under $20.00. The numbers of deaths or births were allocated to these five economic groups on the basis of the residence address of the deceased or of the mother of the newborn child.

DIFFERENTIALS IN MORTALITY FROM ALL CAUSES

The general differences in age-specific mortality rates were found to be similar to those established in the literature. The following conclusions may be drawn from an examination of these rates:

(1) Generally the death rate for females for the city as a whole is lower than the death rate for males. This is true for native whites and Negroes and for all groups with the exception of the foreign born in certain economic groups (Table 1).

(2) There is an inverse relationship between socio-economic status and mortality rates standardized for age, for both sexes, for Buffalo, 1940. (Table 1).

(3) Of the total deaths in Buffalo in 1940 (Table 2), 19.1 per cent would not have taken place if all economic groups had the same age-specific death rates as the highest economic group (general index). Death "wastage" due to socio-economic differences was greater for

TABLE 1

TOTAL MORTALITY RATES BY SEX, NATIVITY, COLOR AND SOCIO-ECONOMIC GROUPS, BUFFALO, 1940; RATES STANDARDIZED FOR AGE

(Per 1,000)

Socio-Economic Group	Total Population	Total White	Native White	Foreign White	Negro*
			Male		
All groups	11.7	11 0	11 0	11.1	18.0
I (high)	9 4	9.4	9 4	9 4	—
II	9.9	9.9	10.0	8.7	—
III	11.4	11 4	11.4	11 1	—
IV	12.2	12 2	12 8	10 7	—
V	14 9	14 3	13 8	14 3	18 5
			Female		
All groups	10 1	9.9	9 3	11.3	15.4
I (high)	7.2	7.2	6 7	8 7	—
II	8.1	8.1	7.6	9 5	—
III	9 0	9.0	8.6	12 3	—
IV	10 0	10.0	9 6	10.1	—
V	12.4	11.7	10.8	12 3	15.9

* In this and subsequent tables, in addition to total rates for Negroes, only Group V rates for Negroes are presented inasmuch as the numbers in the other groups were too small to yield stable rates.

females than for males. The proportion of deaths which would not have occurred if there were no mortality differences due to economic status increased inversely with the status group (simple indexes).

(4) Life expectation at birth for both sexes was inversely related to economic status (Table 3). In comparable ages, the life expectation of native whites, both male and female, was higher than that of foreign whites. The life expectation at birth for Negro males of the lowest economic group was more than six years lower than that of native white males of the same economic group and more than twelve years lower than that of native white males of the highest group. Expectation of life at birth for Negro females of the lowest economic group was eight years lower than that of native white females of the same group and about fifteen years lower than that of native white females of the highest economic group.

CAUSE-SPECIFIC MORTALITY DIFFERENTIALS

The foregoing analysis of mortality rates by socio-economic groups was confined to data on deaths from all causes It would be of great value to investigate which specific causes of death are more or less "sensitive" to socio-economic differentiation. Cause-specific mortality analysis from a sociological point of view constitutes a method of refining general differences and helps clarify some of the more intangible socio-psychological and biological factors involved in demographic phenomena

The cause-specific analysis to follow is based on the assumption that

TABLE 2

SIMPLE AND GENERAL INDEXES OF DIFFERENTIAL MORTALITY; BY SEX, BUFFALO, 1940

(Per 1000)

Economic Group	Total Population	Male	Female
	General Indexes		
All Groups	19.1	17 6	20.8
	Simple Indexes		
I (high)	0.0	0.0	0 0
II	6.8	2.3	11.1
III	17 3	15.9	18.8
IV	24.5	21.4	28.0
V	38.5	36 7	41.7

the clinical diagnosis of the cause of death as reported on death certificates is accurate. Considerable literature [7] indicates that significant errors are involved in such diagnoses. All conclusions in this study must, therefore, be considered with this in mind. In addition, because of the relatively small number of deaths, the multiple cross-classification of the study has in some instances resulted in small frequencies. This introduces a serious problem of instability in the resulting rates. However, analysis of mortality data in the study revealed patterns similar to those with more stable rates established in the literature.

The selected causes of death [10] have shown an over-all relationship with economic status differentiation (Tables 4 and 5). In proceeding with the study of the nature and significance of this relationship, it must be stressed that any interpretation is somewhat impaired by the multifactorial determinant of mortality. Lack of early diagnosis, exposure, inadequacy of nutrition, lack of therapeutic knowledge and other cultural and socio-economic factors are found for different causes of death to be important in differing degrees.

TABLE 3

LIFE EXPECTATION AT BIRTH AND AGE 30; BY SEX, RACE, NATIVITY AND ECONOMIC GROUP, BUFFALO, 1940

Economic Group	Total Population		Native White		Foreign White	Negro	
	At Birth	Age 30	At Birth	Age 30	Age 30	At Birth	Age 30
				Males			
All groups	62.9	37.6	63.5	38.2	37.5	54.5	30.8
I (high)	65.7	39.8	66.1	40.1	39.0	—	—
II	65.5	39.9	65.3	39.7	40.5	—	—
III	63.4	38.1	63.6	38.2	37.8	—	—
IV	62.2	37.1	61.6	36.5	38.3	—	—
V	58.2	34.1	59.7	35.4	35.0	53.6	30.4
				Females			
All groups	66.1	40.5	65.0	39.1	38.3	56.9	34.8
I (high)	69.6	43.3	70.3	44.4	40.6	—	—
II	68.3	41.8	68.8	42.4	40.8	—	—
III	66.4	40.7	66.8	41.1	40.0	—	—
IV	64.8	39.1	65.6	39.8	38.7	—	—
V	61.8	37.1	64.1	38.8	36.9	55.8	34.1

The magnitude of the general indexes of differential mortality indicates that economic differences in deaths from tuberculosis, respiratory diseases, heart diseases and intracranial lesions are inversely related to socio-economic status. The rank ordering of the observed differentials coincided also with the degree of complexity of the cause of deaths as measured by the reduction of mortality from such diseases experienced within the past decades. On the basis of the recorded resident deaths

TABLE 4

SIMPLE AND GENERAL INDEXES* OF DIFFERENTIAL MORTALITY; BY SEX AND SPECIFIC CAUSES OF DEATH, BUFFALO, 1940

(Per 1,000)

Economic Group	All Causes†	Tuberculosis	Cancer	Diabetes Mellitus	Intracranial Lesions	Heart Diseases	Respiratory Diseases	Nephritis	All Other Causes
General Index—Total Population									
All Groups	19.1	59.3	10.0	31.9	13.3	13.6	26.2	7.6	26.3
Simple Index—Total Population									
II	6.8	25.7	7.9	23.5	2.9	2.1	−3.9	11.1	9.9
III	17.3	52.6	8.5	34.4	12.0	11.6	23.0	18.8	22.5
IV	24.5	62.3	10.8	41.5	18.2	17.1	36.5	19.1	31.1
V	38.5	81.4	21.5	43.2	31.1	33.3	50.7	−58.8	47.5
General Index—Male									
All Groups	17.6	67.7	17.9	−7.6	7.3	10.7	22.9	−2.2	19.8
Simple Index—Male									
II	2.3	40.9	8.1	−17.6	0.0	0.0	−18.5	4.3	3.2
III	15.9	61.8	19.2	4.3	7.1	9.1	16.7	11.1	20.9
IV	21.4	69.2	23.3	−13.3	4.5	9.9	36.2	9.3	29.7
V	36.4	84.8	29.1	−15.4	21.7	28.8	46.8	−100.0	46.8
General Index—Female									
All Groups	20.8	44.8	2.6	51.9	18.3	17.0	31.1	16.2	27.8
Simple Index—Female									
II	11.1	0.0	7.8	44.1	5.2	4.6	12.5	16.7	17.1
III	18.8	39.1	−1.8	51.2	15.9	14.4	31.2	25.4	24.4
IV	28.0	50.0	−3.9	63.2	29.1	25.7	37.0	28.3	33.1
V	41.7	74.2	11.1	67.7	40.9	25.4	57.7	−22.2	48.7

* Economic Group I as the basis.

† Includes infant mortality.

between 1900 and 1940 in the State of New York [6], for example, tuberculosis, with the most spectacular state-wide reduction in mortality, shows in this study the highest economic difference (Table 6).

Economic differentials were less pronounced in mortality from heart diseases, intracranial lesions, cancer and nephritis. The observed relationship was curvilinear as compared to the linear relationship for the other causes (Table 4) Such a relationship, barring possible errors

TABLE 5

PROPORTIONAL CAUSE-SPECIFIC INDEXES OF DIFFERENTIAL MORTALITY; BY SEX, BUFFALO, 1940

(Per 1,000)

Cause of Death	Total Population	Male	Female
Tuberculosis	24.8	34.0	14.6
Cancer	13.4	21.9	3.7
Diabetes Mellitus	12.0	−1.8	27.5
Intra-cranial Lesions	10.2	4.8	16.3
Heart Diseases	47.0	38.5	56.6
Respiratory Diseases	12.6	12.3	12.9
Nephritis	4.8	−1.2	11.5
All other causes	66.6	67.9	65.1
Total	191.4	176.4	208.2

TABLE 6

GENERAL INDEXES OF DIFFERENTIAL MORTALITY; BY CAUSE OF DEATH, BUFFALO, 1940

(Per 1,000)

Causes of Death	Total Population	
	Index*	Ratio†
Tuberculosis	59.3	7.8
Diabetes Mellitus	31.9	4.2
All other causes	27.3	3.6
Respiratory Diseases	26.2	3.4
Heart Diseases	13.6	1.8
Intracranial lesions	13.3	1.8
Cancer	10.0	1.3
Nephritis	7.6	1.0

* Index per 1,000 deaths

† Ratio of each index to smallest.

in death certification, might well be due to the economic power of the upper economic group and the available free medical care for the lower group, with the "in-between" groups having to bear their burden with their own means

INFANT MORTALITY DIFFERENTIALS

In spite of medical and social progress, the first year of life still shows the greatest death toll of any single year It is important to locate the areas with need for improvement and to find those biological or socio-economic factors which appear to be associated with mortality during this period.

To offset problems of rate stability caused by the small number of deaths for each of the five economic groups, Groups I and II were combined, Group III remained intact, and Groups IV and V were combined.

The following may be stated on the basis of the study's findings:

(1) Mortality rate of infants, neonatal and one to eleven months old, is greater for females than for males (Table 7)

(2) There is an inverse relationship between infant mortality and economic status. Economic differences are present even during the first day after birth (for males) and become more pronounced as age at death of infants increases (Table 7)

TABLE 7

INFANT MORTALITY RATES BY SEX, RACE, AGE AND ECONOMIC GROUPS, BUFFALO, 1940

(Per 100,000)

Economic Group and Sex	Under 1 day		1–29 days		Under 1 month		1–11 months		Total under 1 year	
	White	Negro	White	Negro	White	Negro	White	Negro	White	Negro
Groups I and II										
Male	1519		1291		2810		630		3440	
Female	1309		855		2164		611†		2775	
Group III										
Male	1624		1624		3248		1052		4300	
Female	1103		1164		2267		1042		3309	
Groups IV and V										
Male	1574	1916†	1778	1342†	3352	3258†	1590	1724*	4942	4982
Female	1136	2565†	1134	1099†	2270	3664†	1325	1139*	3595	4803
All Groups										
Male	1568	1916†	1591	1341†	3159	3257†	1174	1727*	4338	4984
Female	1178	2565†	1094	1092†	2237	3663†	1043	1098*	3280	4761

* Less than 10 instances of death.

† Between 10 and 24 deaths.

TABLE 8

NEONATAL AND INFANT DEATHS AND MORTALITY RATES BY SEX, RACE AND SPECIFIC CAUSE; BUFFALO, 1940

(Rates per 100,000)

Race	Total Under 1 year				Under 1 month				1-11 months			
	Male		Female		Male		Female		Male		Female	
	Deaths	Rates	Deaths	Rates	Deaths	Rates	Deaths	Rates	Deaths	Rates	Deaths	Rates
All Causes												
Total	598	4350	443*	3349	434	3157	303	2291	164	1193	140*	1058
White	572	4333	415	3280	417	3159	283	2237	155	1174	132	1043
Negro	26	4981	26	4762	17	3257	20	3663	9	1724	6	1099
Respiratory Diseases												
Total	77	560	56*	423	22	160	10	76	55	400	46*	348
White	72	545	50	395	21	160	10	79	51	385	40	316
Negro	5	958	4	733	1	192	—	—	4	766	4	733
Diarrhea												
Total	36	262	19	144	8	58	7	53	28	204	12	91
White	36	273	19	150	8	61	7	55	28	212	12	95
Negro	—	—	—	—	—	—	—	—	—	—	—	—
Congenital Malformations												
Total	74	538	72	544	51	371	45	340	23	167	27	204
White	74	560	71	561	51	386	44	348	23	175	27	213
Negro	—	—	1	183	—	—	1	183	—	—	—	—
Premature Birth												
Total	207	1506	145	1096	204	1484	142	1074	3	22	3	22
White	191	1447	129	1020	188	1424	126	996	3	22	3	24
Negro	16	3065	16	2931	16	3065	16	2931	—	—	—	—
Injury at Birth												
Total	59	429	32	242	58	422	32	242	1	7	—	—
White	59	447	31	245	58	439	31	245	1	8	—	—
Negro	—	—	1	183	—	—	1	183	—	—	—	—
Other Diseases Peculiar to First Year of Life												
Total	67	487	52	393	63	458	49	370	4	29	3	22
White	67	508	50	395	63	477	47	372	4	31	3	24
Negro	—	—	2	366	—	—	2	366	—	—	—	—
All Other Causes												
Total	78	568	67	507	28	204	18	136	50	364	49	371
White	73	553	65	514	28	212	18	142	45	341	47	371
Negro	5	958	2	366	—	—	—	—	5	958	2	366

* Deaths are reported for the 1939–41 period. Two deaths of females of "other races" from respiratory diseases were not designated as such in this tabulation.

TABLE 9

TOTAL INFANT MORTALITY RATES BY SEX, RACE AND ECONOMIC GROUP BY SELECTED CAUSES OF DEATH; BUFFALO, 1940

(Rates per 100,000)

Economic Group	All Causes		Diseases of the Respiratory System		Diarrhea		Congenital Malformations	
	Male	Female	Male	Female	Male	Female	Male	Female
White								
I	3623	2865	362*	92*	91*	92*	725*	832*
II	3374	2743	387†	338*	35*	150*	492†	564†
III	4340	3310	443†	429†	236*	61*	591†	368†
IV	4493	3413	595†	471†	352†	235*	514†	677†
V	5668	3870	922†	445†	599†	178*	599†	534†
Negro								
V	5495	5252	1099*	840*	—	—	—	210*

Economic Group	Premature Birth		Injury at Birth		Other Diseases Peculiar to First Year of Life		All Other Causes	
	Male	Female	Male	Female	Male	Female	Male	Female
White								
I	1087†	1017†	544*	462*	453*	185*	362*	185*
II	1160	902	422†	263*	422†	263*	457†	263*
III	1594	1165	413†	245*	561†	398†	502†	644†
IV	1407	1000	460†	177*	487†	441†	677†	412†
V	1843	979†	461†	222*	599†	578†	645†	934†
Negro								
V	3297*	3151*	—	210*	—	420*	1099*	420*

* Less than 10 instances of death.
† Between 10 and 24 deaths.

(3) Economic differentials among males were more pronounced than among females in neonatal mortality (Table 7).

(4) The incidence of Negro infants' deaths was greater than that of white infants.

(5) The most frequent causes of mortality at an age less than one month were premature birth, injury at birth, and congenital malformations. Mortality from premature birth and from congenital malformations has not been reduced as drastically as that from other causes. It was mortality from these two causes which showed no significant relationship with economic group (Tables 8 and 9)

(6) For infant mortality after the first month of life, premature birth, injury at birth and congenital malformations gave way to other causes such as respiratory diseases and diarrhea. General mortality from such diseases has been drastically reduced during the past 50 years. In this study, it was mortality from these causes that was inversely related to economic status.

It will be difficult to state precisely all those factors which influence mortality for this age group. Yerushalmy [13] has indicated that variations in neonatal mortality are dependent upon congenital differences to a larger degree than are variations in adult or infant mortality as a whole. He shows that mortality varies with the age of the mother and order of birth and that there is a greater possibility for the first born to experience trauma and certain conditions of congenital malformations If these results are relevant in the present study, it may be that such congenital factors as order of birth, age of parents, etc , offset economic inequalities. This hypothesis would to a certain extent explain the absence of any relationship between economic status and mortality from congenital malformations and permature birth.

Similarly, the existence of some direct relationship between economic status and mortality from injury at birth could be explained by the hypothesis that there was in the period 1939–41 a direct relationship between the proportion of first order births and economic status and that consequently the liability of a first born injury was directly related to economic status Another hypothesis is that the better the economic status of the mother, the more likely is the use of instrumentation in childbirth as contrasted with the more prevalent use of spontaneous childbirth by the lower status groups, and that consequently there is a greater probability of injury at birth for the better economic status groups.

SUMMARY

Socio-economic status, as measured by median monthly rentals, was shown to be inversely related to mortality in the city of Buffalo for 1939–41. Such a relationship was reflected in striking differences in expectation of life among various groups, and in the proportion of "wasted" deaths due to differences in mortality rates between the highest economic group and the others. The inverse relationship was present even for mortality of infants under one day and was more pronounced for deaths at greater ages.

Cause-specific mortality analysis demonstrated that mortality due to selected causes, which have been shown to be related to environmental conditions and for which medical improvements have reduced the mortality considerably, exhibited the greatest economic differentials On the other hand, mortality from premature birth and congenital malformations, among the most frequent biological causes of neonatal mortality, showed no relationship.

These analyses for the city of Buffalo in 1940 have further substantiated hypotheses previously tested in other cities. This study has further illustrated the usefulness of rentals as a device for socio-economic status differentiation and has demonstrated once more that demographic phenomena are highly sensitive to socio-economic differentiation.

REFERENCES

[1] Convis, Lologene, *Economic differentials in causes of death*. Unpublished A.M thesis, Department of Sociology, University of Chicago, 1939, pp 10 ff

[2] Dublin, Louis I., Lotka, Alfred J , and Spiegelman, Mortimer, *Length of Life*. New York: The Ronald Press Co., 1949, pp 50 ff.; 121 ff.; 212 ff

[3] Hauser, Philip M , *Differential fertility, mortality and reproduction in Chicago, 1930*. Unpublished Ph.D thesis, Department of Sociology, University of Chicago, 1938, pp. 34 ff

[4] Kitagawa, Evelyn M., "Differential fertility in Chicago, 1920–40," *American Journal of Sociology*, LVIII (March, 1953) 481–92.

[5] Mayer, Albert J. and Hauser, Philip M , "Class differentials in expectation of life at birth," *Revue de l'Institute International de Statistique*, 18, 197–200

[6] New York, State of, *Sixty-fifth Annual Report of the Department of Health*, vol. 2, 1945. 1–234.

[7] Pohlen, Kurt and Emerson, Haven, "Errors in clinical statements of causes of death," *American Journal of Public Health*, XXXII (1941–42) 251–60 and XXXIII (1942–43) 505–16

[8] Terris, Milton, "Relation of economic status to tuberculosis mortality, by age and sex," *American Journal of Public Health*, XXXVIII (1948) 1061–70

[9] Titmuss, Richard M., *Birth, Poverty and Wealth*, London: Hamish-Hamilton Medical Books, 1943, 36 ff.

[10] U. S. Department of Commerce, Bureau of Census, *Manual of the International List of Causes of Death as Adopted for Use in the United States*, 1939 revision, Washington: Government Printing Office, 1946, 13–52.

[11] Westergaard, Harold, "Scope and method in statistics," *Publications* of the *American Statistical Association*, XV (1916–17), 225–7.

[12] Yeracaris, Constantine A., *Differential mortality, general and cause-specific, Buffalo, 1940*. Unpublished Ph.D thesis, Department of Sociology, University of Chicago, 1953, 1–174.

[13] Yerushalmy, Jacob, "The effect of order of birth and age of parents upon neonatal mortality," *American Journal of Hygiene*, XXVIII (1938) 244–70.

INTEREST AS A SOURCE OF PERSONAL INCOME AND TAX REVENUE*

A Report of the National Bureau of Economic Research

LAWRENCE H. SELTZER
Wayne University

CONTENTS

INTRODUCTION

CHANGES in interest rates in recent years have sometimes given rise to heated controversies about their economic effects and the public policies involved. The present study was not undertaken because of any special concern with such controversies but as part of a larger analysis of income taxation in the United States. Nevertheless, some of the facts revealed may be useful to those who debate policy.

For example, it is relevant to know something about the distribution of interest income in discussing debt management or tax policy. A once popular belief was that the bulk of interest goes to a class of wealthy investors who are enabled to live by clipping bond interest coupons and receiving interest on mortgages rather than by more direct contributions to production. How far this assumption is from the present situation in the United States is indicated by some of the figures contained in this report It will be found, for example, that the importance of interest as a source of personal income has shrunk strikingly in recent years and that much the greater part of the total interest income re-

* This paper arose out of a study, still in process, of the personal income tax. I am indebted to various members of the staff of the National Bureau of Economic Research and collaborators for criticism and suggestions, notably to W. L. Crum, Solomon Fabricant, Raymond W. Goldsmith, Daniel M Holland, C. Harry Kahn, John W. Kendrick, Geoffrey H. Moore, George Soule, and, for valuable statistical assistance, to Arnold Oliphant and Bella Shapiro I am also indebted to Charles F Schwartz and Selma F. Goldsmith of the Department of Commerce, and Marius Farioletti of the Bureau of Internal Revenue, for supplying special tabulations or making useful suggestions. The charts were drawn by H. Irving Forman.

This paper has been approved for publication, as a report of the National Bureau of Economic Research, by the Director of Research and the Board of Directors of the National Bureau, in accordance with the resolution of the Board governing National Bureau reports (see the 35th Annual Report of the National Bureau of Economic Research, May 1955). It is to be reprinted as No. 51 in the National Bureau's series of Occasional Papers, and the Occasional Paper may be ordered from the National Bureau of Economic Research, 261 Madison Avenue, New York 16, N. Y.

ceived by individuals is received by those with small and moderate incomes.

Another subject of debate is the influence of tax policy on the choice of investments, that is, on the inducement to invest in relatively safe interest-bearing securities as against possibly risky equities, or vice versa. This study offers some revealing evidence in this connection, indicating, among other things, the great importance of the prevailing expectations with respect to the possibilities of capital gains.

Is the tendency to higher interest rates which has been in evidence since 1946, despite some interruptions and reversals, likely to continue, and, if so, are interest rates likely to rise to former peaks? On these questions, too, some light is thrown by analysis of the statistics and related considerations.

It will doubtless surprise many persons to learn that only about one-fourth of the interest income estimated to have been received by individuals in the years 1930–1950 was reported on taxable income tax returns, despite the sharp decreases during the 1940's in the level of personal exemptions and the huge increase in the number of taxable returns filed. What are the important reasons for this fact?

It is the aim of the National Bureau of Economic Research, in this as in its other studies, not to take a position on disputed issues of public policy but rather to make available empirical evidence and impartial analysis for whatever uses may be appropriate.

The present paper begins by tracing the changing absolute and relative importance of personal interest receipts and their changing distribution as between higher and lower income groups. Then it examines the principal causes of the long decline in personal interest income after 1929 and attempts to account, so far as possible, for the fact that only about one-fourth of total personal interest receipts have been reported on taxable individual income tax returns in recent years. The third section of the paper presents some measures of the importance of interest income as a source of tax revenues. The fourth discusses the effects of the progressive income tax rate structure upon interest yields and interest income after taxes and the resulting influence upon incentives to individuals to invest in taxable fixed-interest securities. A final section considers some probable future trends in personal interest income.

I. THE SHIFTING IMPORTANCE OF INTEREST AS A SOURCE OF PERSONAL INCOME

1. *Substantial fluctuations in the absolute and relative importance of interest have occurred since 1913*

Although interest has the reputation of being a relatively stable

TABLE 1

INTEREST COMPONENT OF PERSONAL INCOME RECEIPTS, 1913–1953

Year	Total income receipts of individuals (billions of dollars)	Monetary interest	Interest as percentage of total income receipts	Year	Total income receipts of individuals (billions of dollars)	Monetary interest	Interest as percentage of total income receipts
1913	32.5	1.4	4.4	1932	45.6	4.6	10.2
1914	31.9	1.5	4.6	1933	44.2	4.4	10.0
1915	34.1	1.6	4.6	1934	50.3	4.2	8.4
1916	40.7	1.6	4.0	1935	56.9	4.0	7.1
1917	49.5	1.8	3.6	1936	65.1	3.9	6.0
1918	55.2	2.0	3.6	1937	70.2	3.9	5.5
1919	63.1	2.6	4.1	1938	64.6	3.8	5.9
				1939	69.1	3.8	5.4
1919	63.7	2.9	4.6	1940	74.8	3.7	5.0
1920	66.9	3.3	5.0	1941	92.6	3.7	4.0
1921	53.3	3.5	6.6	1942	118.9	3.5	2.9
1922	57.3	3.6	6.3	1943	145.3	3.4	2.3
1923	66.5	3.8	5.7	1944	137.4	3.5	2.2
1924	66.9	3.9	5.8	1945	162.0	3.9	2.4
1925	70.8	4.1	5.7	1946	172.5	4.4	2.5
1926	73.7	4.1	5.6	1947	185.2	4.9	2.6
1927	74.1	4.3	5.8	1948	201.6	5.3	2.6
1928	75.9	4.6	6.0	1949	197.8	5.7	2.9
1929	80.2	4.8	6.0	1950	219.2	6.2	2.8
				1951	245.8	6.8	2.8
1929	80.1	5.2	6.5	1952	260.1	7.0	2.7
1930	71.0	4.8	6.8	1953	274.3	7.5	2.8
1931	60.3	4.8	8.0				

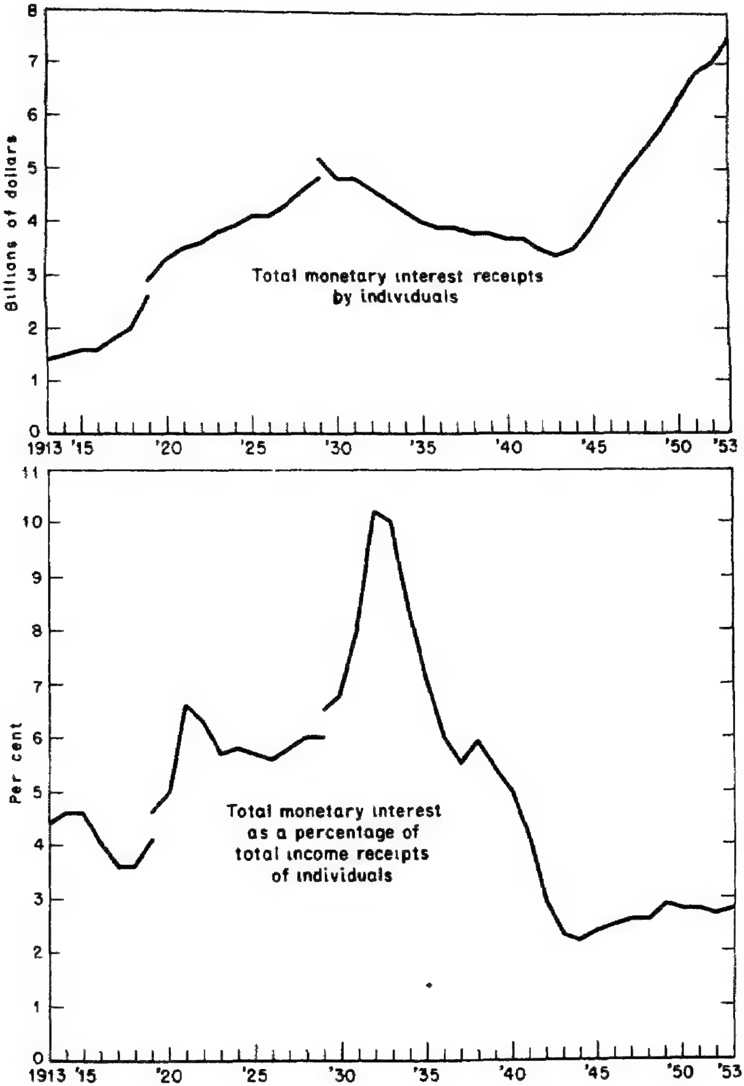
Source Simon Kuznets, *Shares of Upper Income Groups in Income and Savings* (National Bureau of Economic Research, 1953) pp. 570–571, 576–577. Kuznets uses three separate series: (a) 1913–1919 estimates of W. I. King, *The National Income and Its Purchasing Power* (NBER, 1930), and an unpublished extension of the 1913 data, modified by Kuznets (Simon Kuznets, *National Product in Wartime* [NBER, 1945]); (b) 1919–1929, Simon Kuznets, *National Income and Its Composition, 1919–1938* (NBER, 1941). (c) 1929–1953, *National Income Supplement*, 1954, *Survey of Current Business;* and *Statistical Abstract of the United States, 1955.*

Excluded from the figures used here for total personal income are imputed rent, imputed interest inventory valuation adjustments, business transfers except cash prizes, and government contributions to military family allowances.

Interest receipts of nonprofit institutions furnishing services to individuals and current accruals of interest (discount) on individuals' holdings of United States savings bonds are included, while interest received by individuals from other individuals are excluded, from the personal interest income figures of the Department of Commerce which are used here.

CHART 1

INTEREST COMPONENT OF PERSONAL INCOME RECEIPTS, 1913–1953



Source Table 1

source of income, the annual amounts received by individuals in the United States have varied substantially in both directions since 1913. This may be seen in Table 1 and Chart 1, which present annual estimates of total explicit or monetary interest receipts of individuals in the United States in the period 1913–1953. The figures were obtained by joining estimates made by Simon Kuznets for 1913–1929 with those of the Department of Commerce for years beginning with 1929.[1]

Annual receipts of interest by individuals more than tripled between 1913 and 1929. Then the total fell during the next fourteen years, reaching a level in 1943 about 35 per cent below that of 1929. In contrast, dividends paid to individuals in 1943 were only 23 per cent lower than in 1929, while employee compensation, entrepreneurial income, and rents were all much higher.[2] Personal interest receipts began a vigorous upturn in 1944, but despite a great growth in public and private debt they did not overtake the 1929 total until 1948. A comparison of the year-to-year movements of personal receipts of interest and dividends in 1913–1953 is shown in Table 2 and Chart 2.

Most striking in the record of personal interest receipts is the long and pronounced decline in their relative importance as a component of total personal incomes since 1933 (Table 1 and lower section of Chart 1). Interest had contributed a rising proportion of total personal income receipts between 1913 and 1929. In 1929, the first year of the present series of national income estimates of the Department of Commerce, the money interest receipts of individuals constituted 6.5 per cent of their total income receipts. During the early years of the Great Depression the more drastic shrinkage of other sources of income caused the share of interest in total personal income to rise above 10 per cent in 1932 and 1933, though the absolute amount fell substantially. After 1933 the percentage importance of interest declined sharply and almost uninterruptedly until it reached a low of slightly more than 2 per cent in 1944. Since then the proportion has fluctuated somewhat between 2½ and 3 per cent.

The record low levels in the relative importance of interest in total personal income receipts in recent years resulted in largest part from the vastly disproportionate increases in other sources of income. Whereas employee compensation rose from $46 billion to $200 billion between 1939 and 1953 and entrepreneurial income, from $11 billion to $38 billion, money interest receipts of individuals rose only from $3.8 billion to $7.5 billion.[3]

---

[1] See note to Table 1.

[2] *National Income Supplement, 1954, Survey of Current Business*, Dept. of Commerce.

[3] *Ibid.*, and special tabulation by Dept. of Commerce of net rent of owner-occupied farm dwellings, the amounts of which we deduct from entrepreneurial income.

2. Interest income more widely diffused among the different income groups

The figures for 1917–1948 presented by Simon Kuznets in his *Shares of Upper Income Groups in Income and Savings* and those compiled from income tax returns for subsequent years show that a marked reduction has taken place in the concentration of interest receipts in the upper income groups. Correspondingly, there has been an increased diffusion of interest income among the population as a whole. Whereas the top 5 per cent of the population, when ranked according to the size of family income per member, received 51 per cent of the total personal interest income in 1919 and an average of 43.4 per cent of the total in 1918–1938, their proportion had fallen to 26 per cent in 1948.[4] The share of the top 1 per cent of the population fell from 37 per cent of all personal interest income in 1919 to 15¾ per cent in 1948. Conversely, the lower 95 per cent of the population, who had obtained only 49 per cent of total personal interest receipts in 1919, received 74 per cent in 1948

Although the relative importance of interest as a component of personal income declined substantially for all income groups, the decline was greatest in the upper income groups. In 1919 the top 1 per cent of the population obtained 13 4 per cent of their aggregate income from interest; ten years later the proportion was still 12 9 per cent; but by 1948 it had shrunk to 4.6 per cent.[5] The top 5 per cent of the population, whose interest receipts had accounted for 10.3 per cent of their total income in 1919 and 10 9 per cent in 1929, obtained only 3 7 per cent of their total income from this source in 1948 For the lower 95 per cent of the population the percentage of income obtained from interest was 2.9 per cent in 1919 and 2 2 per cent in 1948.

These figures, as well as those from income tax returns in recent years, are in contrast with a still common notion that members of the uppermost income groups in the United States obtain a large fraction of their incomes by clipping and cashing maturing interest coupons on their bond holdings and from interest receipts on mortgages and other loans. The income tax figures indicate that in the ten years ending in 1950, taxable interest contributed an average of only about 3 cents of each dollar of adjusted gross income reported by all taxpayers with incomes exceeding $25,000. (Adjusted gross income is statutory net income before the allowable personal deductions such as interest and taxes paid, charitable contributions, medical expenses, and casualty losses.) The major sources of taxable individual incomes over $25,000

[4] Simon Kuznets, *Shares of Upper Income Groups in Income and Savings* (National Bureau of Economic Research, 1953), Table 123, pp. 647 ff.

[5] *Ibid.*, Table 125, pp 668 ff.

in 1941–1950 were dividends, salaries, and entrepreneurial income from unincorporated business. In 1950 the proportion that taxable interest contributed to the total adjusted gross income of each of nine income size groups was as follows:

Income size group	Per cent
$500,000 and over	2.9
$100,000–$500,000	3.0
50,000– 100,000	2.9
25,000– 50,000	2.8
10,000– 25,000	2.2
5,000– 10,000	0.9
3,000– 5,000	0.4
2,000– 3,000	0.5
Less than $ 2,000	0.8

TABLE 2

COMPARISON OF INTEREST AND DIVIDENDS IN PERSONAL INCOME, 1913–1953

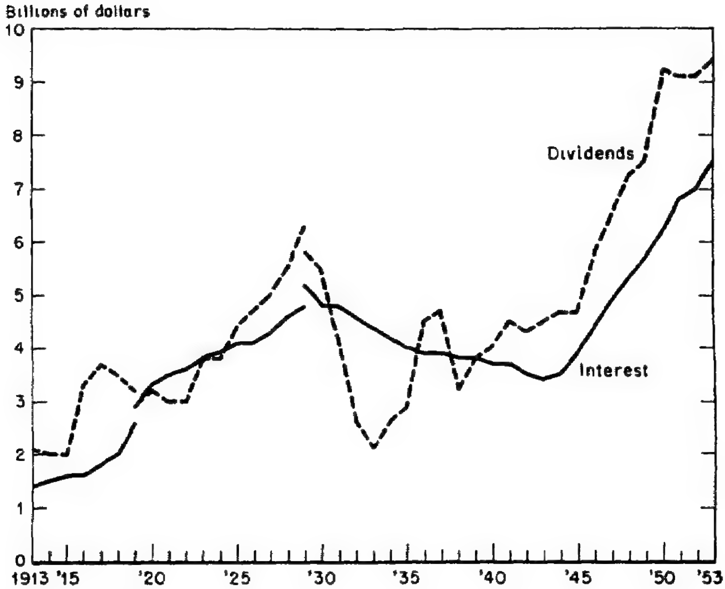
(Billions of Dollars)

Year	Dividends	Monetary interest	Year	Dividends	Monetary interest
1913	2.1	1.4	1932	2.6	4.6
1914	2.0	1.5	1933	2.1	4.4
1915	2.0	1.6	1934	2.6	4.2
1916	3.3	1.6	1935	2.9	4.0
1917	3.7	1.8	1936	4.5	3.9
1918	3.5	2.0	1937	4.7	3.9
1919	3.2	2.6	1938	3.2	3.8
			1939	3.8	3.8
1919	2.9	2.9	1940	4.0	3.7
1920	3.2	3.3	1941	4.5	3.7
1921	3.0	3.5	1942	4.3	3.5
1922	3.0	3.6	1943	4.5	3.4
1923	3.8	3.8	1944	4.7	3.5
1924	3.8	3.9	1945	4.7	3.9
1925	4.4	4.1	1946	5.8	4.4
1926	4.7	4.1	1947	6.5	4.9
1927	5.0	4.3	1948	7.2	5.3
1928	5.5	4.6	1949	7.5	5.7
1929	6.3	4.8	1950	9.2	6.2
			1951	9.1	6.8
1929	5.8	5.2	1952	9.1	7.0
1930	5.5	4.8	1953	9.4	7.5
1931	4.1	4.8			

Source. Same as in Table 1.

CHART 2

COMPARISON OF INTEREST AND DIVIDENDS IN PERSONAL INCOME, 1913–1953



Source Table 2.

Of the 4,410,271 tax returns reporting interest income in 1950, 3,667,607 of them reported interest income of $500 or less, and only 27,161 reported interest income of more than $5,000 (Table 3). The latter group accounted for 15 8 per cent of the total interest income reported on all returns

Unlike Kuznets' estimates the figures drawn from income tax returns are subject to the qualification that they do not include interest from tax-exempt securities. Such securities are doubtless held in relatively greater amounts among the upper income groups than among the lower. Nevertheless, as recently as 1941—the latest year for which this information was reported on income tax returns—individuals with incomes of $5,000 and over reported only 39 per cent of the state and local government securities estimated to be in the hands of individuals.[6] After allowing for the holdings of taxable fiduciaries and for underreporting, George Lent's estimates indicate that individuals with

[6] George E. Lent, *The Ownership of Tax-Exempt Securities, 1913–1953* (National Bureau of Economic Research, Occasional Paper 47 (1954), Table 7 Between 1924 and 1942 the law required individuals with incon of $5,000 or more to re ort on their income tax returns the amount of their holdin

TABLE 3

ESTIMATED INTEREST ON ALL RETURNS REPORTING INTEREST, BY SIZE OF INTEREST INCOME REPORTED, 1950

Size of interest income reported (thousands of dollars)	No. of returns in each group (1)	Average interest reported per return (2)	Total interest income of each group (1)×(2) (thousands of dollars) (3)	Percentage of total interest in each group (3)÷total (per cent) (4)
Under 0.1	2,041,369	50	102,068	6.17
0 1–0.2	782,152	150	117,323	7.09
0.2–0 3	406,642	250	101,660	6.14
0.3–0.4	267,209	350	93,523	5.65
0.4–0 5	170,235	450	76,606	4.63
0.5–1.0	409,048	750	306,786	18.53
1.0–1 5	138,969	1,250	173,711	10.49
1 5–2 0	65,106	1,750	113,936	6.88
2 0–2.5	37,753	2,250	84,944	5.13
2.5–3 0	22,858	2,750	62,860	3.80
3–4	26,829	3,500	93,902	5.67
4–5	14,940	4,500	67,230	4.06
5–10	20,114	6,385.6	128,440	7.76
10–25	6,054	14,564.5	88,173	5.33
25–30	809	33,735 7	27,292	1.65
50–100	148	66,878.0	9,898	0.60
101 or over	36	198,775 9	7,156	0.43
Total	4,410,271		1,655,508	100.01
Actual total reported in *Statistics of Income*			1,595,604	
Overestimate			59,904	
Per cent overestimated			3 75	

Source: *For returns reporting interest income of less than $5,000, the average amount of interest per return in each interest size group was estimated by assuming that it equalled the midpoint of the size group. The number of returns in each group was obtained from* Statistics of Income *Because of the wider limits of the interest size groups of $5,000 and over, the midpoint was not deemed representative of the average. In these groups the average amount of interest per return was estimated by assuming that the number of returns within each interest size group, as reported in* Statistics of Income, *was distributed similarly to the distribution of returns reporting adjusted gross income of the same amounts.*

incomes of less than $5,000 owned nearly one-half of the state and local government securities in individual hands in 1941.[7] We estimate below,

[7] Unpublished estimate of George E. Lent.

moreover, that individuals' interest receipts from tax-exempt securities fell by more than one-half between 1932 and 1949.

In the taxable estate tax returns filed in 1949, as reported in *Statistics of Income for 1948*, wholly tax-exempt securities constituted 0.5 per cent of gross estates under $100,000, 1 per cent of those between $100,000 and $500,000, 3.2 per cent of those between $500,000 and $1 million, 8.4 per cent of $1 million to $5 million, 17 per cent of $5 million to $10 million, and 32.2 per cent of gross estates of $10 million and over. For estates under $5 million the proportions represented by tax-exempt securities in 1949 were considerably smaller, on the average, than the proportions in the years 1934–1945 and for estates of $5 million and over, slightly greater.

Although the exclusion of tax-exempt interest causes the role of interest in the composition of income to be somewhat understated for all income groups in figures taken from income tax returns, and to be more understated for larger than for smaller incomes, the understatement is relatively small for any broad income size group. Tax-exempt interest doubtless bulks large in the incomes of some individuals. But entrepreneurial profits, dividends, capital gains, and salaries dwarf tax-exempt interest in importance for each group as a whole This is true even of the highest income groups, which owe the bulk of their aggregate incomes to dividends, entrepreneurial profits, and capital gains.[8] Even if, in the face of the sharp decline in total tax-exempt interest received by individuals and of the evidence from estate tax returns, we were to assume that the amount of fully tax-exempt interest received in 1949 by taxpayers with adjusted gross income of $100,000 or more was twice as great as the amount of such interest reported by these income groups in 1940, their tax-exempt interest would equal only 3.9 per cent of the adjusted gross income they reported Thomas R. Atkinson found that in 1949 state and local government obligations represented only 2.3 per cent of the total financial assets of all Wisconsin individuals with incomes of $50,000 or more and only 1.2 per cent of the financial assets of those with incomes of $20,000–50,000.[9]

The lower among taxable income groups have received a surprisingly large porportion of the aggregate interest income reported on taxable individual returns in recent years. Individuals with incomes under $5,000 accounted for an average of 41 per cent of the annual totals of taxable interest in the ten years 1941–1950, and even those with incomes under $3,000 accounted for an average of 25 per cent of the annual totals in this period (Tables 4 and 5).

[8] Kuznets, *op. cit.*, Table 125, pp 668 ff.

[9] Thomas R. Atkinson, "The Pattern of Financial Asset Ownership" (National Bureau of Economic Research, mimeographed).

The lower among the taxable income groups have consistently received a much larger proportion of total interest income than of total dividend receipts reported on taxable returns. Taxpayers with incomes under $5,000 accounted for an average of only 13 per cent of the annual total of dividends reported on taxable returns in the 33 years 1918–1950 as compared with 36 per cent of the total interest. In 1941–1950 they accounted for an average of 21 per cent of the total dividends so reported as against 41 per cent of total interest.

These figures accord in a general way with Atkinson's findings with respect to the pattern of financial asset ownership of individuals in Wisconsin in 1949[10] Atkinson found that individuals with incomes less than $5,000 tend to concentrate much of their financial assets in debt obligations, particularly such kinds as savings bank deposits, building and loan association shares, and life insurance policies but that corporate stocks become progressively more important in higher income groups Similarly, Joseph A. Pechman, in analyzing the patterns of income revealed by a study of the 1936 Wisconsin state income tax returns, found that interest was the second most frequent source of income for wage earners with less than $2,000 income and for entrepreneurs with incomes under $1,000, as well as for entrepreneurs with incomes between $6,000 and $8,000[11]

3. *Long decline in personal interest income after 1929 and relatively moderate growth in recent years in the face of large increases in debt*

It is easy to point to the major reasons for the substantial decline in the absolute amounts of interest received by individuals during the 1930's but less easy to account fully for the only moderate recovery in the 1940's, when immense increases took place in the amount of interest-bearing debt outstanding in the United States.

In the years immediately following 1929 the shrinkage in personal interest income reflected in part a general contraction in the total amount of interest payments, both to individuals and corporations. Such payments fell sharply in the early years of the Great Depression as bank and other loans were reduced and as defaults took place on large amounts of bonds of railroad and other domestic corporations, foreign governments and corporations, and on residential and farm mortgages. Gross annual monetary interest payments in the United States fell from $13.6 billion in 1929 to $8.8 billion in 1935, as estimated

[10] *Ibid.*

[11] Frank A. Hanna, Joseph A Pechman, and Sidney M. Lerner, *Analysis of Wisconsin Income*, Studies in Income and Wealth, Volume Nine (National Bureau of Economic Research, 1948), Chap 2.

TABLE 4

PERCENTAGE SHARES OF DIFFERENT INCOME GROUPS IN TOTAL INTEREST REPORTED ON TAXABLE INDIVIDUAL AND FIDUCIARY RETURNS, 1918–1950

	Net income classes* (thousands of dollars)								
	Less than 2	2–3	3–5	5–10	10–25	25–50	50–100	100–500	500 and Over
1918	7.7	12.7	18.5	16.9	15.3	9.2	7.1	8.7	4.0
1919	6.9	8.2	17.5	18.3	18.2	10.9	7.8	9.1	2.9
1920	7.5	12.6	18.6	18.1	18.5	10.5	7.1	5.5	1.5
1921	8.8	9.2	20.8	19.5	19.7	10.6	6.3	4.3	0.8
1922	7.3	8.0	19.5	19.3	20.0	11.5	7.5	5.5	1.5
1923	9.7	10.9	27.5	15.3	16.7	9.2	5.5	4.1	1.1
1924	9.0	10.8	29.4	13.0	16.2	9.5	6.1	4.7	1.3
1925	4.1	6.3	18.2	17.5	22.7	13.4	8.5	7.0	2.3
1926	4.0	6.3	15.2	19.1	23.7	13.3	8.7	7.3	2.3
1927	3.3	5.7	13.1	19.6	24.6	13.8	9.2	8.0	2.7
1928	3.4	6.4	13.4	18.6	22.6	13.1	9.3	9.0	4.2
1929	2.9	5.7	12.4	18.8	23.5	13.4	9.3	9.3	4.8
1930	4.7	7.4	14.4	20.8	24.8	12.5	7.6	5.9	2.1
1931	4.0	7.4	14.3	23.6	27.2	11.4	6.6	4.1	1.4
1932	9.6	9.3	17.4	22.2	20.5	11.5	5.8	3.2	0.6
1933	11.2	9.9	17.2	21.8	19.4	10.7	5.6	3.5	0.7
1934	7.6	6.8	15.6	24.3	25.1	11.5	5.5	2.9	0.7
1935	6.9	6.5	15.0	23.7	25.1	12.3	6.3	3.5	0.7
1936	9.8	8.6	17.7	19.1	20.7	11.3	6.9	4.7	1.3
1937	10.0	8.1	17.4	20.1	21.3	10.9	6.4	4.5	1.3
1938	15.6	9.0	18.3	20.5	19.9	8.6	4.1	3.0	1.0
1939	15.2	8.8	17.5	19.8	19.9	9.4	4.7	3.5	1.2
1940	18.6	13.4	17.3	17.1	17.1	8.2	4.1	3.1	1.0
1941	23.5	16.3	18.1	14.1	14.1	6.7	3.6	2.9	0.8
1942	25.7	15.0	15.2	13.7	15.1	7.7	4.1	2.9	0.5
1943	23.7	12.5	14.4	15.4	16.4	8.6	4.9	3.6	0.6
1944	15.7	10.9	16.0	17.6	18.7	9.7	5.9	4.2	1.4
1945	13.9	10.9	14.6	18.1	20.0	10.4	6.2	4.5	1.4
1946	10.9	9.9	14.9	18.2	21.1	11.4	6.8	5.2	1.6
1947	10.9	10.1	14.9	18.8	20.9	11.1	6.4	5.3	1.6
1948	5.4	7.6	15.7	21.0	21.9	12.6	7.8	6.5	1.7
1949	6.6	8.4	17.2	22.1	20.6	11.5	6.9	5.4	1.4
1950	5.3	6.9	15.0	21.5	22.0	13.4	7.9	6.3	1.7

* In 1944 changed to Adjusted Gross Income Classes.

Note. The figures do not necessarily add to 100 per cent due to rounding.

Source: *Statistics of Income*. For the years 1918 through 1930, the figures for interest include annuities and dividends from foreign corporations (except those with more than 50 per cent of gross income from sources in the United States), both of which were lumped with interest as "Interest and other income" in *Statistics of Income*.

For [illegible] *of I* [illegible]

TABLE 5

INTEREST REPORTED ON TAXABLE RETURNS, BY INCOME CLASSES, 1918–1950

(Millions of Dollars)

Year	Net income classes* (thousands of dollars)									Total
	Less than 2	2–3	3–5	5–10	10–25	25–50	50–100	100–500	Over 500	
1918	99.9	165.5	241.0	220.6	199.5	120.5	92.9	113.8	51.8	1,305.6
1919	96.7	114.9	244.4	255.5	254.6	152.3	108.6	127.2	41.1	1,395.3
1920	111.2	186.3	276.2	268.7	274.8	156.0	105.2	81.4	22.3	1,482.0
1921	110.5	115.8	263.0	246.4	249.3	134.4	79.5	53.8	9.7	1,262.2
1922	104.7	115.4	281.2	278.8	289.0	165.3	107.7	78.9	21.4	1,441.8
1923	181.8	204.0	514.1	285.3	311.5	172.7	103.5	76.3	20.2	1,869.3
1924	181.6	216.1	591.4	260.4	326.3	190.6	122.7	94.7	25.2	2,009.0
1925	67.3	103.3	297.9	287.7	372.4	219.7	140.3	114.7	37.8	1,641.0
1926	70.5	111.2	267.9	336.1	416.9	233.1	152.1	128.5	40.8	1,757.1
1927	59.7	104.6	240.2	359.0	451.4	252.2	169.4	147.4	49.7	1,833.6
1928	69.7	130.2	272.8	377.7	459.5	265.3	188.5	183.6	85.2	2,032.5
1929	58.0	114.1	249.7	380.3	473.4	270.5	188.1	187.4	97.0	2,018.5
1930	79.5	126.3	245.7	354.6	423.1	213.2	130.4	100.3	35.4	1,708.4
1931	46.2	84.1	162.8	269.1	310.4	130.0	75.7	47.1	15.6	1,141.0
1932	96.7	93.9	176.0	224.6	207.0	116.2	59.2	32.1	6.4	1,012.1
1933	94.7	83.9	145.9	184.7	164.6	90.8	47.3	29.5	6.2	847.5
1934	61.8	55.1	127.3	197.5	204.5	93.9	44.4	23.5	6.0	813.9
1935	56.5	53.1	121.6	193.0	204.1	99.7	51.5	28.2	5.3	813.0
1936	88.5	77.8	159.8	172.0	186.8	101.8	62.4	42.3	11.3	902.7
1937	88.4	71.8	153.1	177.1	187.9	96.1	56.7	39.5	11.6	882.3
1938	135.3	77.7	158.7	177.2	172.4	74.8	35.2	25.7	9.1	866.1
1939	136.5	78.9	157.6	178.3	179.2	84.9	42.1	31.5	10.9	900.3
1940	183.9	132.6	170.8	169.0	169.2	80.8	40.4	30.5	10.3	987.5
1941	269.7	187.0	207.5	161.8	161.6	76.9	40.9	33.8	9.0	1,148.2
1942	293.5	171.9	173.5	157.1	172.7	87.9	46.3	33.3	6.2	1,142.3
1943	259.2	136.5	158.2	168.2	179.2	94.5	63.4	39.3	7.0	1,095.4
1944	169.1	118.2	172.5	190.4	201.4	104.2	63.3	46.9	14.7	1,079.7
1945	148.8	116.1	155.8	193.3	214.2	111.4	66.7	47.6	14.8	1,068.6
1946	133.7	122.0	182.4	222.9	259.5	139.4	83.7	64.3	19.7	1,227.6
1947	136.2	126.1	185.5	234.2	260.2	138.7	80.2	65.5	19.9	1,246.5
1948	70.0	98.4	204.0	273.2	284.4	164.0	101.1	84.5	21.5	1,301.2
1949	99.8	126.7	260.6	335.1	311.7	173.6	104.0	81.8	21.5	1,514.8
1950	86.1	111.6	243.1	348.7	356.0	216.3	127.1	102.1	27.5	1,618.5

* In 1944 changed to Adjusted Gross Income Classes.

Note: The figures do not necessarily add to total due to rounding.

Source: Same as in Table 4.

nontaxable returns with taxable returns for net incomes of \$5,000–10,000 and \$5,000–6,000, interest on taxable returns was estimated by assuming that it constituted the same ratio of total reported interest as the ratio of net income on taxable returns to total net income reported.

For the years 1941–1943, interest was combined with dividends, rents, royalties (for 1941 only), and annuities on 1040A (short-form) returns. Interest was estimated for these years by assuming it constituted the same ratio of these components of income as in 1040 (long-form) returns.

by the Department of Commerce. Total personal interest receipts fell from \$5.2 billion in 1929 to \$4.0 billion in 1935. The amount of personal interest income continued to shrink until the entrance of the United States into World War II and did not rise above its 1929 level until 1948 despite a large and rapid growth in the total of public and private debt during the war and postwar years.

The net amounts of public and private debt outstanding, as defined by the Department of Commerce but modified to exclude debt bearing no interest, are shown for selected years, by classes of borrowers, in Table 6. Net public and private debt is defined to exclude government obligations held by federal, state, or local governments or their agencies or trust funds; corporate obligations held by the issuer or its close affiliate; and short-term debts of individuals and nonfinancial unincorporated business enterprises owed to other such enterprises and individuals. It also excludes bank deposits and the value of outstanding life insurance and annuity policies. The total fell from \$160 billion in 1930 to \$142 billion in 1933. It rose slowly during the remainder of the 1930's, approximating its 1930 level in 1940 Then it more than doubled in the next five years, reaching \$362 billion in 1945. The whole of this \$202 billion increase in this period was in the federal debt and was mainly occasioned by World War II. The federal government then reduced its debt by \$24 billion between 1945 and 1953, but the net debt of other borrowers more than doubled, causing the total net public and private interest-bearing debt at the end of 1953 to be greater than that at the end of 1945 by \$141 billion, or 39 per cent, and greater than that at the end of 1929 by \$345 billion, or 218 per cent.

For 1943, interest includes partially tax-exempt dividends.

For 1944–1945, when interest and dividends were reported together on income tax returns, interest was estimated by assuming that the ratios between interest and dividends in 1944–1945 were the same as in 1946.

Prior to 1937, interest received by fiduciaries was not tabulated separately from their other income. We estimated interest to constitute the following percentages of the adjusted gross income of fiduciaries in this period. 1922–1930, 0 33, 1931–1935, 0 40, 1936, same as 1937 For years beginning with 1937, the amount of interest included in individuals' taxable income from fiduciaries was estimated to constitute the same proportion of their total income from fiduciaries as interest was of the adjusted gross income of fiduciaries.

Two elements in these procedures work to produce small opposing errors (1) Some understatement of interest occurs because long-term capital gains are included in the adjusted gross income of fiduciaries, thereby making the ratio of interest to their adjusted gross income lower than it otherwise would be. Since long-term capital gains are reported separately by the beneficiaries and are not included in income from fiduciaries, our estimate of interest from fiduciaries is somewhat less than it should be. (2) Some overstatement of interest arises, on the other hand, because a portion of interest directly reported on taxable fiduciary returns is distributed to beneficiaries as income from fiduciaries and is thus included twice in our estimates of interest The duplication appears to be very small, however. In 1947, for example, if all the interest income of fiduciaries went to beneficiaries filing taxable returns—an extreme assumption, the maximum overestimation of interest from fiduciaries would be about 3 per cent.

Excluded from our figures for 1944–1947 and 1948–1950 is the interest portion of the total of dividends and interest not exceeding \$100 per return, reported by taxpayers filing W2 returns in the former period and 1040A in the latter.

TABLE 6

NET INTEREST-BEARING PUBLIC AND PRIVATE DEBT AT END OF CALENDAR YEARS, 1918–1953

(Billions of Dollars)

	Federal government	State and local government	Long-term corporate	Farm mortgage	Nonfarm mortgage	Short-term corporate	Non-mortgage, non-corporate	Total interest-bearing debt
1918	21	5	30	7	10	7	27	106
1919	25	5	31	8	10	7	24	112
1920	23	6	33	10	12	8	25	117
1921	23	7	34	11	13	8	25	120
1922	23	8	34	11	14	9	25	123
1923	21	8	36	11	16	9	25	127
1924	21	9	39	10	19	9	26	132
1925	20	10	40	10	21	10	27	137
1926	19	11	42	10	24	10	27	142
1927	18	11	44	10	27	10	27	148
1928	17	12	46	10	30	10	29	154
1929	16	13	47	10	31	11	29	158
1930	16	14	51	9	32	9	28	160
1931	18	16	50	9	31	8	23	154
1932	21	16	49	8	29	6	18	148
1933	24	16	48	8	26	5	15	142
1934	30	16	45	8	26	5	15	143
1935	33	16	44	7	25	5	16	145
1936	37	16	43	7	24	5	17	149
1937	39	16	44	7	24	7	18	154
1938	40	16	45	7	24	5	17	154
1939	42	16	44	7	25	5	17	156
1940	44	16	44	6	26	5	18	160
1941	56	16	44	6	27	6	20	175
1942	101	15	43	6	27	8	15	214
1943	153	14	41	5	26	8	15	263
1944	210	14	40	5	26	9	17	321
1945	250	13	38	5	27	7	20	362
1946	228	14	41	5	32	9	20	350
1947	221	15	46	5	39	12	24	361
1948	214	17	52	5	45	12	29	376
1949	216	19	57	6	51	12	33	394
1950	216	22	60	6	59	18	39	420
1951	216	24	66	7	67	22	42	444
1952	221	27	73	7	75	22	48	474
1953	226	30	79	8	84	23	53	503

Sources. Federal debt: *Survey of Current Business*, September 1953, October 1954; *Annual Report of the Secretary of the Treasury*, 1952 Matured and non-interest-bearing debt was deducted from the total net federal debt outstanding at the end of calendar years as reported in the *Survey of Current Business* for years beginning with 1929; for prior years the June 30 figures for matured and non-interest-bearing debt as reported in the Treasury's annual report were averaged to obtain December 31 esti-

Certain important sources of interest income of individuals are excluded from the figures in Table 6 Among these are savings and time deposits in commercial and savings banks, the total amount of which increased from $22.1 billion to $63.6 billion between the end of 1924 and of 1952. Except for mortgage debt the amount of interest-bearing obligations of individuals and unincorporated enterprises owed to noncorporate creditors is also excluded.

4. *Decline in interest rates*

Contributing most conspicuously to the contraction of total interest payments after 1932 was the prolonged decline in interest rates that then began. One type of interest rate was reduced to zero and a related one severely limited by the Banking Act of 1933 which prohibited the payment of interest on demand deposits by any bank insured by the Federal Deposit Insurance Corporation and provided for regulating the maximum rate of interest that could be offered by such banks on time and savings deposits

Interest rates on new issues of corporate and governmental debt securities and market yields on old issues fell almost continuously between 1932 and 1946 During this period the monthly average yield on Moody's Aaa corporate bonds fell from 5.01 per cent in 1932 to 2 53 per cent in 1946, that on Moody's Baa bonds fell from 9.30 to 3.05 per cent; and the average yield of long-term Treasury bonds fell from 3.68 for bonds issued when the interest was exempt from normal tax to 2.19 per cent for fully taxable bonds (Table 7, Chart 3). (For corporate investors the latter yield was reduced to about 1.66 per cent by the normal tax of 24 per cent, and to still lower levels for corporations with

---

mates, and these were deducted from the December 31 totals of net federal debt reported in the *Survey of Current Business*.

State and local government debt. *Survey of Current Business*, September 1953, October 1954, May 1955, *Indebtedness in the United States*, Dept. of Commerce, 1942; *Governmental Debt in 1952*, Dept. of Commerce, 1952. Non-interest-bearing debt was deducted from the Commerce figures, the amounts for 1929–1941 and 1944–1952 being obtained from the Commerce volumes cited, and estimated for the other years. The June 30 interest-bearing debt figures were averaged to obtain December 31 estimates.

Long-Term Corporate, Farm Mortgage, and Non-Farm Mortgage, *Survey of Current Business*, September 1953, October 1954.

Short-Term Corporate R. W Goldsmith's unpublished NBER monograph on capital requirements, *Statistics of Income* Figures for 1912, 1922, 1929, and 1933 represent the non-real estate loans of all operating commercial banks to corporations as reported by Goldsmith Figures for the intervening years were obtained by straight-line interpolation Those for 1934–1936 are interpolations of Goldsmith's figures for 1933 and 1939. Figures for 1937–1952 were estimated from the balance sheet data in *Statistics of Income*, adjusted to cover corporations not submitting balance sheets.

Nonmortgage-Noncorporate *Survey of Current Business*, September 1953, October 1954; Rolf Nugent, *Consumer Credit and Economic Stability* (Russell Sage Foundation, 1939), *Federal Reserve Bulletin*, April 1953. From the total net nonmortgage-noncorporate debt figures reported by the Dept. of Commerce, we deducted our estimates of consumers' non-interest-bearing charge account and service debt, estimates based on the data reported in Nugent, *op cit.*, and the *Federal Reserve Bulletin*.

TABLE 7

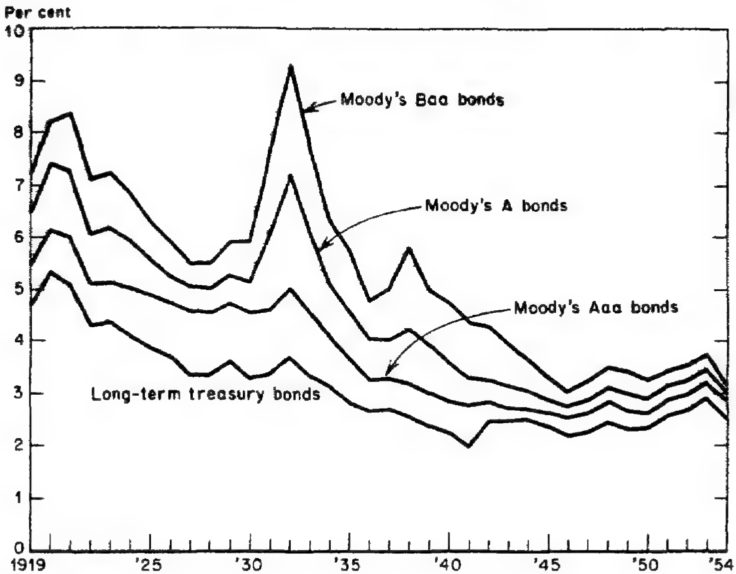
YIELDS ON UNITED STATES TREASURY AND CORPORATE BONDS, 1919–1954

Year	Long-term treasury bonds	Moody's corporate bonds		
		Aaa	A	Baa
1919	4.73	5.49	6.48	7.25
1920	5.32	6.12	7.41	8.20
1921	5.09	5.97	7.28	8.35
1922	4.30	5.10	6.03	7.08
1923	4.36	5.12	6.17	7.24
1924	4.06	5.00	5.93	6.83
1925	3.86	4.88	5.55	6.27
1926	3.68	4.73	5.24	5.87
1927	3.34	4.57	5.04	5.48
1928	3.33	4.55	5.01	5.48
1929	3.60	4.73	5.28	5.90
1930	3.29	4.55	5.13	5.90
1931	3.34	4.58	6.01	7.62
1932	3.68	5.01	7.20	9.30
1933	3.31	4.49	6.09	7.76
1934	3.12	4.00	5.08	6.32
1935	2.79	3.60	4.55	5.75
1936	2.65	3.24	4.02	4.77
1937	2.68	3.26	4.01	5.03
1938	2.56	3.19	4.22	5.80
1939	2.36	3.01	3.89	4.96
1940	2.21	2.84	3.57	4.75
1941	1.95	2.77	3.30	4.33
1942	2.46	2.83	3.28	4.28
1943	2.47	2.73	3.13	3.91
1944	2.48	2.72	3.06	3.61
1945	2.37	2.62	2.87	3.29
1946	2.19	2.53	2.75	3.05
1947	2.25	2.61	2.87	3.24
1948	2.44	2.82	3.12	3.47
1949	2.31	2.66	3.00	3.42
1950	2.32	2.62	2.89	3.24
1951	2.57	2.86	3.13	3.41
1952	2.68	2.96	3.23	3.52
1953	2.93	3.20	3.47	3.73
1954	2.54	2.90	3.06	3.18

Source: *Treasury Bulletin* and Moody's Investors Service as reprinted in *Survey of Current Business* Between Jan. 1, 1919 and Oct. 15, 1925, the Long-Term Treasury Bond series included all outstanding partially tax-exempt bonds not due or callable for 8 years or more, between Oct 15, 1925, and Dec. 31, 1941, those not due or callable for 12 years or more; between the latter date and April 1, 1952, all fully taxable issues not due or callable for 15 years or more; since the last date, all fully taxable issues not due or callable for 12 years or more.

CHART 3

YIELDS ON UNITED STATES TREASURY AND CORPORATE BONDS, 1919–1954



Source Table 7

incomes subject to surtax) Many millions of dollars of high-grade public utility and other corporate bonds, as well as much larger amounts of United States Treasury bonds, were refunded at, or in advance of, maturity during the latter half of the 1930's and in the 1940's by new issues carrying lower interest rates

Reflecting reduced interest rates, primarily, five of the seven major sources of *private* interest payments, exclusive of financial institutions—all but consumer credit and nonfarm mortgages—were smaller in 1948 than in 1929, as may be seen in the figures for selected years in Table 8. Total interest payments from these seven sources, which had been $7.9 billion in 1929, fell to $4.6 billion in 1940, after which they began an ascent which brought them to $5 6 billion in 1948 and to $10.4 billion in 1953.

5. *Institutionalization of investment*

One reason for the reduced level of total personal interest income in

TABLE 8

MONETARY INTEREST PAYMENTS ON SELECTED TYPES OF DEBT, 1929–1953

(Millions of Dollars)

	Mortgages on nonfarm dwellings	Farm mortgages	Debt of transportation and utility corporations	Debt of other non-banking corporations	Consumer credit	Loans from brokers	Net inflow of interest from abroad	Total interest, selected types of private debt	Government interest payments
1929	1,560	582	1,387	2,358	443	951	577	7,858	1,506
1930	1,549	570	1,468	2,312	444	264	608	7,215	1,513
1931	1,571	553	1,567	2,023	366	100	550	6,730	1,521
1932	1,538	526	1,551	1,768	259	45	426	6,113	1,574
1933	1,425	472	1,490	1,524	196	32	324	5,463	1,689
1934	1,336	430	1,314	1,691	221	32	242	5,266	1,849
1935	1,275	396	1,293	1,588	276	15	207	5,050	1,831
1936	1,232	364	1,244	1,500	386	33	195	4,954	1,868
1937	1,214	341	1,183	1,498	490	38	160	4,924	2,019
1938	1,192	320	1,176	1,382	458	22	138	4,688	1,920
1939	1,167	305	1,156	1,336	500	25	127	4,616	1,941
1940	1,172	293	1,077	1,345	616	20	120	4,643	2,059
1941	1,217	284	1,035	1,320	743	17	126	4,742	2,068
1942	1,238	272	1,053	1,207	530	15	130	4,445	2,407
1943	1,184	246	1,035	1,105	326	20	115	4,031	3,141
1944	1,147	230	978	1,097	303	26	118	3,899	3,895
1945	1,131	220	951	1,095	328	32	130	3,887	4,934
1946	1,204	216	838	1,122	496	25	135	4,036	5,772
1947	1,372	222	839	1,340	784	18	168	4,743	5,751
1948	1,592	229	873	1,548	1,085	23	224	5,574	5,904
1949	1,783	242	959	1,735	1,310	26	230	6,285	6,196
1950	2,061	264	996	1,857	1,675	48	248	7,149	6,428
1951	2,361	291	1,063	2,228	1,857	65	312	8,177	6,652
1952	2,649	319	1,149	2,464	2,110	72	317	9,080	7,023
1953	2,974	349	1,220	2,732	2,644	90	333	10,351	7,441

Source. Office of Business Economics, Dept. of Commerce and *National Income Supplement, 1954 Survey of Current Business*

spite of the immense increase in the total outstanding public and private debt (Table 6) has been the continuing growth of financial intermediaries as holders of interest-yielding securities. Raymond W. Goldsmith has estimated that the main groups of financial intermediaries, excluding personal trust departments of banks but including government lending institutions and social security systems, increased their total assets by $257 billion, or more than 206 per cent, between 1929 and 1949 (Table 9).[12] They raised their share of the total owner-

[12] Raymond W. Goldsmith, *The Share of Financial Intermediaries in National Wealth and National Assets, 1900–1949* (National Bureau of Economic Research, Occasional Paper 42, 1954), Table 1

ship of each of the principal classes of interest-bearing securities outstanding in the United States as follows:[13]

	Per Cent of Total Held by Financial Intermediaries	
	1929	1949
United States government securities	46	64
State and local government securities	34	53
Corporation and foreign bonds	33	76
Mortgages	61	70

In consequence of this development much income that would formerly have been received directly as interest income of individuals now takes the form of additions to various types of claims against financial intermediaries and of dividends from corporations The channeling of savings and investment through private life insurance companies, for example, caused their assets—most of which consist of bonds and mortgages—to increase from less than $9 billion in 1922 to $79 billion in 1953.[14] Interest-yielding investments that many individuals formerly made directly have in more recent years been made indirectly through life insurance companies The effective exemption from personal income tax of the interest that policyholders earn on their accumulating life insurance reserves doubtless influences some persons to make some of their investments in this form rather than in the form of direct ownership of fixed-interest securities

A development that considerably narrowed an important area of individual lending was the widespread adoption, under the leadership of the Federal Housing Administration, of the long-term amortized large value-ratio mortgage in the financing of residential construction and the consequent drastic reduction in the demand for junior lien financing in this field During the 1920's the junior mortgage or subordinated land contract was a large source of interest income for many individuals. Because institutional investors then generally limited their first mortgage loans to 60 per cent or less of the appraised value of the property (the appraisals were sometimes excessively liberal, however), purchasers of houses commonly raised an additional 20 to 25 per cent

[13] *Ibid*, adapted from Tables 11, 13, 14, 16. "Outstanding in the United States" means only the amounts held in this country of the bonds of foreign corporations but the total outstanding issues, whether held in this country or abroad, of American governments, corporations, and mortgages.

[14] *Life Insurance Fact Book*, Institute of Life Insurance.

TABLE 9

TOTAL ASSETS OF MAIN GROUPS OF FINANCIAL INTERMEDIARIES

(Billions of Dollars)

	1929	1949
Banking system:		
Federal Reserve Banks	5.46	45.39
Commercial banks	66 24	157.46
Mutual savings banks	9.87	21.49
Postal Savings System	0.17	3.31
Total	81.74	227.66
Insurance:		
Private life insurance companies	17.48	59.55
Fraternal life insurance organizations	0.85	1.98
Private self-administered pension funds	0 50	6.85
Federal, } pension, retirement, and	0.96	33.98
State and local } social security funds	0.52	4.87
Fire and marine insurance companies	3.08	6.56
Casualty and miscellaneous insurance companies	1.54	5 45
Mutual accident associations	0.09	0.28
Total	25.02	119 53
Miscellaneous financial intermediaries		
Saving and loan associations	7.41	14.55
Credit unions	0.04	0.83
Investment companies	2.94	2 95
Investment-holding companies	4.35	2.44
Investment installment companies	0.05	0 29
Mortgage companies	0.78	0.50
Land banks	1.94	1.01
Government lending institutions	0.29	12.31
Total	17.81	34.87
Grand total	124.57	382 06

Source: Raymond W Goldsmith, *The Share of Financial Intermediaries in National Wealth and National Assets* (National Bureau of Economic Research, Occasional Paper 42, 1954), Table 1. Slight discrepancies between components and totals are due to rounding.

by borrowing from others, largely individuals, on notes secured by a junior lien on the property. Under these circumstances a residential building boom, such as occurred during the 1920's, could lead to a substantial increase in the interest income of individuals. In the building boom that followed World War II, in contrast, institutional investors widely adopted the long-term amortized mortgage—with or without FHA insurance—under which they were often willing to lend 80 per cent or more of the value of the mortgaged property. In consequence, even apart from restrictions by the FHA against junior lien financing in connection with insured mortgages, the area for such financing was greatly reduced.

The direct participation of individual lenders in financing residential construction was further narrowed both by their exclusion from the list of eligible lenders in the FHA program of insured mortgages and by the stimulation of commercial bank lending in this field. Whereas the National Banking Act imposes severe restrictions on the freedom of national banks to make conventional mortgage loans, and substantially similar restrictions exist in many state banking laws, mortgage loans insured by the FHA or guaranteed by the Veterans' Administration were generally exempted from these limitations. Residential mortgage loans by commercial banks were thereby greatly stimulated. In 1926–1930 commercial banks accounted for 14 per cent of home mortgage loans; in 1946–1950 they accounted for 23 per cent of the much larger total [15] Between 1935 and 1950 they originated about one-third of the total amount of FHA-insured home mortgage loans and about one-half of the total amount of FHA-insured loans on rental and cooperative housing projects.[16]

The continued growth of the corporate form of doing business and of owning property likewise tended to reduce direct individual investment in debt obligations. To the extent that ownership of mortgages, corporation and government bonds, and other debt passes from individuals into the hands of property-managing family corporations, financial corporations, and ordinary business enterprises, the interest income becomes merged with other corporate receipts and reaches individuals in the form of dividends and capital gains.

Further evidence of the increased importance of financial institu-

---

[15] See Leo Grebler, *The Role of Federal Credit Aids in Residential Construction* (National Bureau of Economic Research, Occasional Paper 39, 1953), p. 39.

[16] *Ibid.*, p 38.

tions as recipients of interest income is found in Table 10, which compares the amounts of monetary interest (total interest less "imputed" interest as defined below) received by individuals with the amounts received by the principal classes of financial institutions, as estimated by the Department of Commerce. As a group these institutions increased their interest receipts from a total roughly equal to that for individuals in 1929 to one about 20 per cent greater by 1948 and 41 per

TABLE 10

MONETARY INTEREST RECEIVED BY PRINCIPAL TYPES OF FINANCIAL INTERMEDIARIES COMPARED WITH MONETARY INTEREST RECEIVED BY INDIVIDUALS AND NONPROFIT ORGANIZATIONS, 1929–1953

(Millions of Dollars)

	Commercial banks	Federal Reserve banks	Mutual savings banks	Life insurance companies	Corporations in finance, n.e.c	Savings and loan associations	Credit unions	Total for financial intermediaries	Total monetary interest received by individuals and non-profit organizations
1929	2 626	68	516	893	573	548	4	5,228	5,198
1930	2,360	34	523	961	551	539	4	4 972	4,795
1931	2,104	27	550	1,015	374	496	4	4,570	4,843
1932	1,757	48	554	970	338	434	4	4,105	4 645
1933	1,387	48	500	911	275	373	3	3,497	4,421
1934	1,329	48	474	938	483	303	4	3,579	4,239
1935	1,305	41	432	948	492	260	5	3,483	4,039
1936	1,232	37	407	971	486	251	7	3,491	3,898
1937	1,367	40	403	993	522	257	10	3,592	3,883
1938	1,337	35	393	1,030	384	254	12	3,445	3,822
1939	1,369	38	378	1,080	375	365	16	3,621	3,761
1940	1,392	43	377	1,098	425	258	20	3,613	3,701
1941	1,504	41	370	1,129	408	273	25	3,750	3,672
1942	1,487	52	371	1,201	271	285	21	3,688	3,494
1943	1,637	69	375	1,243	214	281	15	3,834	3,391
1944	1,843	104	405	1 293	201	293	14	4,153	3,451
1945	2,083	142	451	1,383	210	317	15	4,601	3,903
1946	2,417	150	503	1,419	272	369	18	5,148	4,376
1947	2,565	158	529	1,502	238	437	27	5,576	4,874
1948	2,800	303	563	1,633	418	549	33	6,299	5,256
1949	2,977	318	606	1,826	461	571	42	6,799	5,737
1950	3,234	278	650	1,970	570	663	58	7,419	6,343
1951	3,626	394	683	2,274	565	750	67	8,359	6,775
1952	4,116	456	757	2,552	633	875	80	9,469	6,980
1953	4,810	513	846	2,782	714	1,051	106	10,822	7,645

Source: Office of Business Economics, Dept. of Commerce and *National Income Supplement, 1954, Survey of Current Business.*

cent greater by 1953. Life insurance carriers approximately doubled their interest receipts between 1929 and 1948, while monetary interest receipts of individuals in 1948 were only slightly above the 1929 level. The amounts for individuals include the interest receipts of nonprofit institutions furnishing services for individuals but exclude interest paid by some individuals to other individuals

Similar evidence is provided by the growth in the absolute amount and relative importance of "imputed interest" as defined and estimated by the Department of Commerce Such interest consists of the amount of the net property income of financial institutions that they spend to provide services for individuals. Table 11 presents figures for 1929–1953 together with those for total personal interest income, which is the sum of the monetary and "imputed" interest receipts of individuals and nonprofit institutions furnishing services to them The imputed interest portion of total personal interest rose from about 30 per cent in 1929 to 44 per cent in 1953, and its absolute amount in the latter year was more than 165 per cent greater than in 1929. The monetary interest of individuals was 45 per cent above its 1929 level in 1953; and its proportion of total personal interest declined from 70 to 56 per cent between 1929 and 1953.

TABLE 11

TOTAL PERSONAL INTEREST, IMPUTED INTEREST PORTION, AND MONETARY INTEREST RECEIVED BY INDIVIDUALS AND NONPROFIT ORGANIZATIONS, 1929–1953

(Billions of Dollars)

Year	Total personal interest	Imputed interest	Monetary interest	Year	Total personal interest	Imputed interest	Monetary interest
1929	7.4	2.2	5.2	1942	5.8	2.3	3.5
1930	6.9	2.2	4.8	1943	5.8	2.4	3.4
1931	6.9	2.1	4.8	1944	6.2	2.7	3.5
1932	6.6	1.9	4.6	1945	6.9	3.0	3.9
1933	6.2	1.8	4.4	1946	7.6	3.2	4.4
1934	6.1	1.9	4.2	1947	8.2	3.3	4.9
1935	5.9	1.9	4.0	1948	9.0	3.7	5.3
1936	5.8	2.0	3.9	1949	9.8	4.0	5.7
1937	5.9	2.0	3.9	1950	10.6	4.4	6.2
1938	5.8	2.0	3.8	1951	11.6	4.8	6.8
1939	5.8	2.0	3.8	1952	12.3	5.3	7.0
1940	5.8	2.1	3.7	1953	13.5	5.9	7.5
1941	5.8	2.2	3.7				

Source *National Income Supplement, 1954, Survey of Current Business*

II. THE SMALL PROPORTION OF PERSONAL INTEREST RECEIPTS REPORTED ON TAXABLE RETURNS

1. *About one-fourth of individuals' interest income reported on taxable returns in the period 1930–1950*

Although they reflect the general movements of total personal interest receipts, the annual amounts of interest reported on taxable individual income tax returns are also importantly influenced by factors other than the movements of total personal interest receipts, such as statutory changes in the level of personal exemptions, changes in the distribution of income, the relative amounts of taxable and tax-exempt interest, etc. It is nevertheless striking to observe that the high point in the absolute amount of interest reported on taxable individual returns was reached as long ago as 1928. In that year, when only 2.5 million taxable income tax returns were filed, the amount of interest reported on them was $2.0 billion. In 1950, when the number of taxable returns was fifteen times as large and the total amount of income on them seven times as large, the amount of interest they included was

TABLE 12

INTEREST INCOME REPORTED ON TAXABLE INDIVIDUAL RETURNS, 1918–1950

	Interest on taxable returns	Adjusted gross income on taxable returns	Interest as percentage of adjusted gross income		Interest on taxable returns	Adjusted gross income on taxable returns	Interest as percentage of adjusted gross income
	(millions of dollars)				(millions of dollars)		
1918	1,305.6	15,545.9	8 4	1935	813.0	11,437 7	7 1
1919	1,395.3	20,064.7	7 0	1936	902 7	16,023 3	5 6
1920	1 482 0	22.893.4	6.5	1937	882.3	17,407.2	5 1
1921	1,262.2	15,667 2	8.1	1938	866 1	14,548 4	6.0
1922	1,441 8	17,348.8	8 3	1939	900 3	17,873 2	5 0
1923	1,869 3	20,295.4	9 2	1940	987 5	26,214 6	3.8
1924	2,009 0	22,207.4	9 0	1941	1,148.2	49,882 0	2.3
1925	1,641.0	19,962 5	8.2	1942	1,142 3	73,216 3	1.6
1926	1,757.1	19,790.5	8.9	1943	1,095 4	105,248 9	1 0
1927	1,833.6	20,322 9	9 0	1944	1,079.7	115,417 0	0.9
1928	2,032.5	23,700 7	8 6	1945	1,068 6	118,418 3	0 9
1929	2,018 5	23,144 2	8 7	1946	1,227.6	119,115.8	1 0
1930	1,708 4	16,011 3	10.7	1947	1,246 5	136,275.5	0 9
1931	1,141 0	10,490 6	10 9	1948	1,301.2	143,043 7	0 9
1932	1,012.1	8,716.1	11.6	1949	1,614 8	139,493 2	1.1
1933	847 5	8,123.4	10 4	1950	1,618.5	159,779.1	1 0
1934	813 9	9,595.3	8 5				

Source: *Statistics of Income* The figures for total adjusted gross income are not perfectly comparable throughout the period. Adjusted gross income is defined in the Internal Revenue Code as gross income

only $1.6 billion, or about 20 per cent less than in 1928 (Tables 12 and 13, Chart 4).

Not only in 1950, but also in each of the preceding nineteen years, the amount of interest reported on taxable individual returns was less than that reported in every year between 1923 and 1930.

Further, whereas interest constituted about 8.6 per cent of the aggregate adjusted gross income of taxable individuals in 1928 and averaged 9.0 per cent of such income annually between 1918 and 1933, its percentage importance had shrunk to 1 per cent by 1943, and has remained at about that level in every year since (up to and including 1950, the latest year for which the figures are available as this is written).

The amounts reported on taxable income tax returns represented only about one-fourth, on the average, of total personal interest receipts in 1930–1950 (Table 13 and Chart 5). Prior to 1930 the proportion was substantially higher.

In 1918, when the low World War I personal exemptions gave wide coverage to the income tax, it had been 66 per cent. After falling steeply to 36 per cent by 1921, it rose to 52 per cent in 1924 and then fluctuated somewhat above 40 per cent during the remainder of the 1920's. Then it plunged to reach a low of 19 per cent in 1933 and 1934. A gradual rise during the next eight years brought the proportion to 33 per cent in 1942, after which it fell off again. It was about 25 per cent in each of the four years 1947–1950.

---

minus allowable trade and business deductions, expenses of travel and lodging in connection with employment, reimbursed expenses in connection with employment, deductions attributable to rents and royalties, deductions for depreciation and depletion allowable to life tenants or to income beneficiaries of property held in trust, and allowable losses from sales of property. Each component of adjusted gross income, therefore, is the net income or loss from that source after making the deductions that are specifically allowed. A net loss from any source of taxable income constitutes a part of the adjusted gross income (or deficit) as well as a net profit (see *Statistics of Income, 1948*, p. 9)

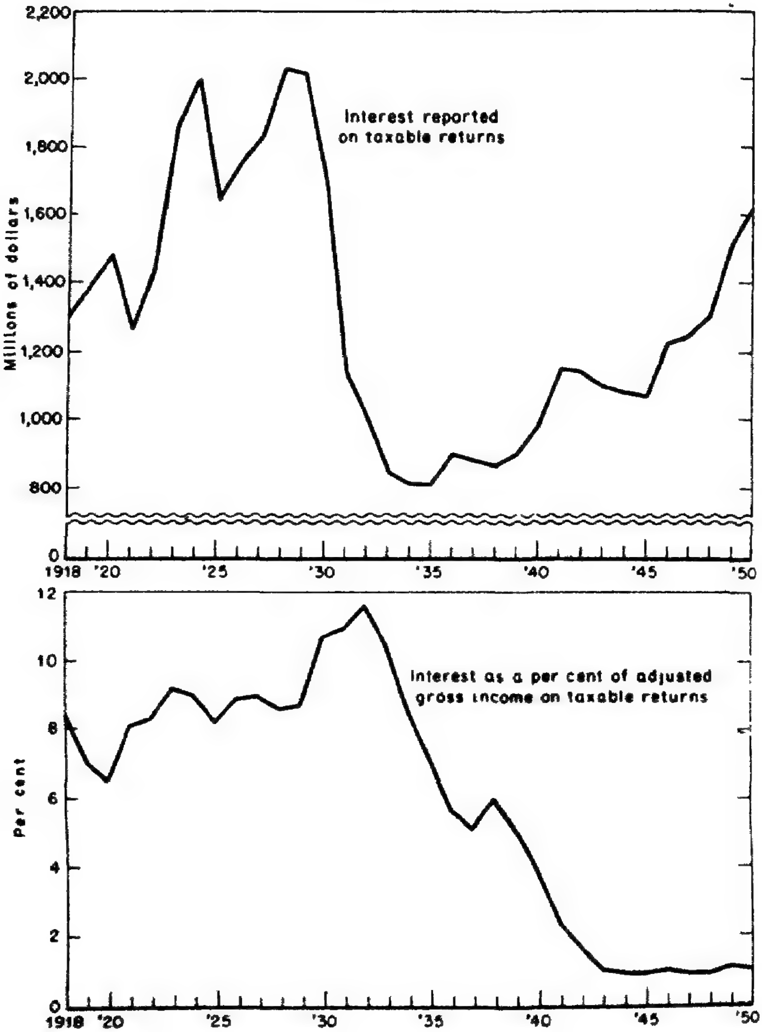
For years prior to 1944, when the Bureau of Internal Revenue first revised its basis of tabulating and reporting from a net income to an adjusted gross income basis, we constructed the adjusted gross income by adding together the separate components as reported in *Statistics of Income*, but had to ignore the amounts of certain negative components of income that were not reported in *Statistics of Income* prior to various years. (1) Prior to 1924, only positive components of income were reported, (2) capital losses from assets held more than two years were first reported in 1924; (3) other capital losses were first reported in 1926, (4) business and partnership losses were first reported in 1930, (5) rent and royalty losses were first reported in 1944, (6) prior year net losses and net operating losses though allowable as a deduction in 1922–1931 and 1940–1943, respectively, were not deducted in our estimates of adjusted gross income for those years.

For the years 1927–1931 and 1932–1936, respectively, when *Statistics of Income* combined nontaxable returns with taxable returns for net incomes of $5,000–10,000, and $5,000–6,000, the various components of income on taxable returns were estimated by assuming they constituted the same ratio of the total reported of the various components as the ratio of net income on taxable returns to total net income reported.

For interest reported on taxable returns, see note to Table 4.

CHART 4

INTEREST AS A COMPONENT OF ADJUSTED GROSS INCOME ON TAXABLE RETURNS, 1918–1950



Source Table 12.

The varying gaps between estimated total individual receipts of interest and the amounts reported on taxable individual income tax returns arise from a number of different causes. Doubtless one of them is that the statistical sources and methods for estimating total interest receipts of individuals are imperfect. The other principal causes are:

a. Changes in the statutory amounts of personal exemptions and credits

b. Fluctuations in national and total personal income

c. Deferred reporting of accrued interest on unredeemed United States savings bonds

d. Changes in tax-exempt interest received by individuals and in total interest received by nonprofit organizations

e Underreporting of interest income for tax purposes

TABLE 13

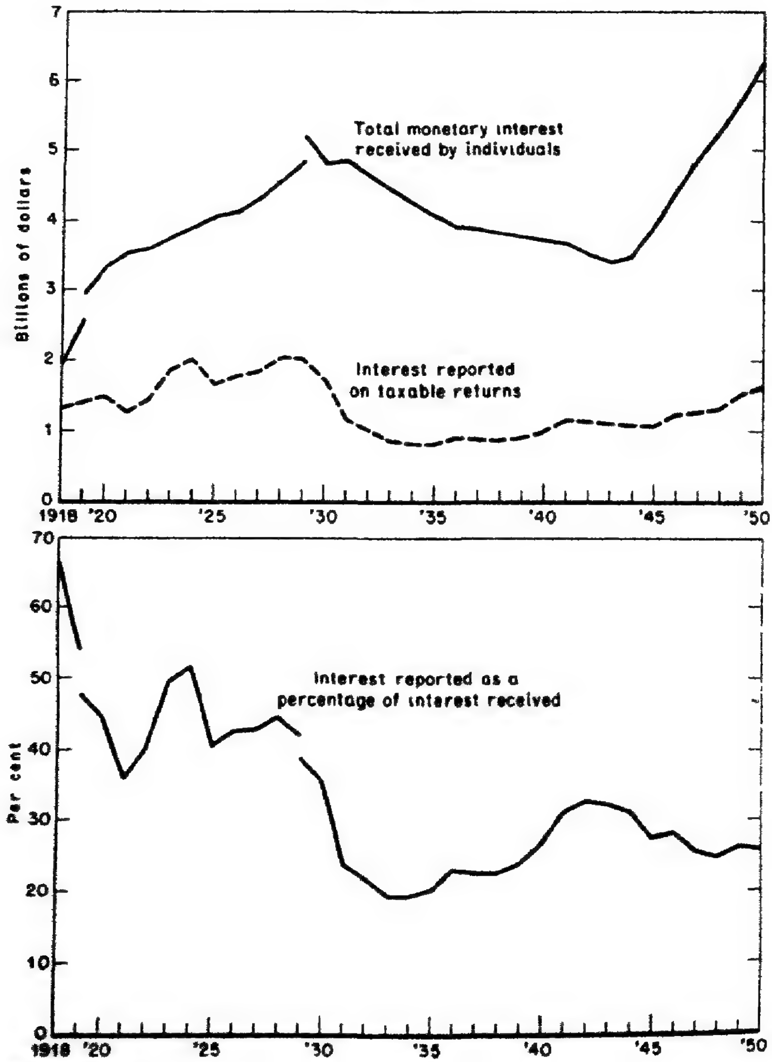
INTEREST REPORTED ON TAXABLE RETURNS COMPARED WITH TOTAL PERSONAL MONETARY INTEREST RECEIPTS 1918–1950

	Total monetary interest	Interest reported on taxable returns	Interest reported on taxable returns as percentage of total monetary interest		Total monetary interest	Interest reported on taxable returns	Interest reported on taxable returns as percentage of total monetary interest
	(billions of dollars)				(billions of dollars)		
1918	1.97	1.31	66.4	1933	4.42	0.85	19.2
1919	2.58	1.40	54.0	1934	4.24	0.81	19.2
				1935	4.04	0.81	20.1
1919	2.93	1.40	47.6	1936	3.89	0.90	23.2
1920	3.33	1.48	44.5	1937	3.88	0.88	22.7
1921	3.52	1.26	35.8	1938	3.82	0.87	22.7
1922	3.59	1.44	40.2	1939	3.76	0.90	23.9
1923	3.77	1.87	49.6	1940	3.70	0.99	26.7
1924	3.90	2.01	51.5	1941	3.67	1.15	31.3
1925	4.06	1.64	40.5	1942	3.49	1.14	32.7
1926	4.12	1.76	42.6	1943	3.39	1.10	32.3
1927	4.30	1.83	42.7	1944	3.45	1.08	31.3
1928	4.57	2.03	44.5	1945	3.90	1.07	27.4
1929	4.83	2.02	41.8	1946	4.38	1.23	28.1
				1947	4.87	1.25	25.6
1929	5.20	2.02	38.8	1948	5.26	1.30	24.8
1930	4.80	1.71	35.6	1949	5.74	1.51	26.4
1931	4.84	1.14	23.6	1950	6.24	1.62	25.9
1932	4.64	1.01	21.8				

Source For monetary interest, Table 1, for interest reported on taxable returns, *Statistics of Income* (see note to Table 4).

CHART 5

INTEREST REPORTED ON TAXABLE RETURNS COMPARED WITH TOTAL PERSONAL MONETARY INTEREST RECEIPTS, 1918–1950



Source: Table 13.

2. *Effects of changes in personal exemptions upon interest reported on taxable returns*

Statutory changes in the amounts of personal exemptions and credits have operated at times to reduce materially, or to add to, the number of taxpayers and the amount of income subject to income tax and, therefore, the total amount of interest on taxable returns. Persons with income under $5,000, which includes those most readily shifted out of or into the taxable category by changes in personal exemptions and credits, generally account for a substantial fraction of all interest reported on taxable returns.

A particularly clear example of the effect on taxable interest of raising personal exemptions may be had by comparing the figures for 1925 with those for 1924. An increase from $2,500 to $3,500 in the personal exemption for heads of families and from $1,000 to $1,500 for single persons was made for 1925 (and continued through 1931) The number of taxable returns filed in 1925 dropped 44 per cent below that of 1924, and their aggregate net income dropped 10 per cent, despite a rise in total national income in 1925 A substantial amount of interest income that had previously been included on taxable returns was now exluded. Interest reported by taxpayers with net incomes of $5,000 or less dropped $521 million below the preceding year despite an increase of $153 million in the amount reported by the other income groups. The larger exemptions and credits also caused a smaller rise than would otherwise have occurred in interest income reported on taxable returns in the later 1920's.

3. *Fluctuations in personal and national income affect the amount of interest reported on taxable returns*

More commonly, the effects of statutory alterations in personal exemptions, though important, have been obscured and overshadowed by the influence of major changes in the level of business activity and national income which have operated to shift large numbers of taxpayers out of or into the taxable category by altering the aggregate amounts of their incomes from all sources. In consequence, taxpayers with unchanged income from interest were often taxable in one year and not in another Deductible capital and other losses, such as were widely incurred in 1921 and 1929–1933, converted many incomes containing substantial interest components into statutory net losses or reduced them below the allowance for personal exemptions and credits for dependents, while in other years some individuals whose interest receipts had previously gone untaxed were brought into the taxable group by increases in their incomes from other sources.

For example, the reduction in taxable interest and total taxable income brought about by the increase in personal exemptions and credits for dependents in 1921 was accentuated by the business recession of that year. In consequence of both the change in law and the recession, the number of taxable returns fell 35 per cent, total statutory net income on them 34 per cent, and the amount of interest included on them 15 per cent below the figures for 1920 With no change in personal exemptions and credits, an extraordinary volume of capital gains, arising largely from the stock market boom, together with a growth of the national income, were major factors in causing the statutory net income of taxable returns in 1925–1928 to rise from $17.5 billion to $21 billion and the amount of interest income reported on them from $1.6 billion to $2.0 billion. Again with no change in personal exemptions the sharp decline in business activity between 1929 and 1931 was accompanied by a decline from $20.5 billion to $9.3 billion in the total net income of taxable returns and from $2.0 billion to $1 1 billion in the amount of interest reported on them. The continuing decline in personal incomes in 1932 outweighed the effects of the reductions in personal exemptions enacted for that year, with the result that total net income and interest income on taxable returns continued to decline

Apart from the low absolute level of interest income reported on taxable individual returns during the 1940's, the small proportion of the reported interest to the total amounts estimated to have been received is particularly noteworthy because many millions of persons and billions of dollars were added to the income tax rolls in this period by the sharply rising levels of total personal income and by reductions in personal exemptions. The number of taxable individual income tax returns, which had ranged between about 2½ and 5½ million during the 1920's, and 1½ to 4 million during the 1930's, rose above 40 million by 1943 and even higher thereafter. In these circumstances it is distinctly surprising that a material increase failed to take place in the proportion of total personal interest receipts accounted for on taxable returns Actually, the proportion in the latter part of the 1940's was little different from that of 1939 and much lower than in the 1920's (Table 13, Chart 5).

4. *United States savings discount bonds*

One factor in the failure of reported taxable interest receipts in the 1940's to recover to earlier levels despite the great increase in private holdings of debt was the large investment made by individuals since 1935 in United States savings bonds, the accruing interest on which

need not be reported for taxation, if the holder so chooses, until the bonds are redeemed. It is probable that little of the accruing interest is reported until redemption and that a substantial fraction is never reported.[17]

Until April 1940, Series A-D bonds were open to purchase by commercial banks and other corporations as well as by individuals. But each purchaser was limited to $7,500 issue price in any one calendar year. This restriction had the effect of largely confining subscriptions to individuals even before corporations were made ineligible as purchasers. Purchases of Series E bonds were open only to natural persons and were limited at first to $3,750 issue price in any calendar year for each purchaser, a limit that was subsequently raised to $10,000 maturity value and in 1952 to $20,000 maturity value. This series of bonds accounted for 61 per cent of the current redemption value of all the unmatured savings bonds outstanding December 31, 1952

Until replaced in May 1952 by similar issues bearing higher yields, Series F and G bonds were continuously open to all subscribers except commercial banks in amounts up to $100,000 issue price per year for the two issues combined ($50,000 in calendar year 1944), and commercial banks were also permitted to purchase these bonds during certain periods and under certain restrictions. Series G bonds, which were not discount bonds unless turned in for redemption in advance of maturity, proved far more popular than Series F and were outstanding in the amount of $18 4 billion current redemption value at the end of 1952

[17] United States savings bonds were first offered for sale in March 1935 Bonds of Series A–D were sold until May 1941, when they were succeeded by Series E, F, and G Series A–E bonds were 10-year discount bonds sold at 75 per cent of maturity value to yield 2 90 per cent per annum if held to maturity and lesser yields if redeemed before maturity Series F bonds were 12-year discount bonds sold at 74 per cent of maturity value to yield 2 53 per cent per annum if held to maturity and less if redeemed prior thereto

Series G bonds were twelve-year current income bonds sold at par and bearing interest at 2½ per cent per annum, but redeemable at the option of the holder in advance of maturity at varying discounts from par. Further details may be found in the *Treasury Bulletin*, December 1944, p 36, and February 1946, p. 37.

A number of changes in the savings bond offerings became effective in May and June 1952 (1) For Series E, the over-all interest rate was raised from 2 9 to 3 per cent without changing the offering price or maturity value, by shortening the maturity from 10 years to 9 years, 8 months, the redemption values were raised for the early and intermediate periods in advance of maturity, and the maximum amount purchasable by any person in any single year was raised to $20,000 maturity value (2) A new Series H current income bond was introduced, available for individuals in amounts up to $20,000 maturity value annually per purchaser, with issue and redemption at par, and with current payment of interest on a graduated scale at rates closely corresponding to the revised E bond scale and the same maturity as the revised E bonds (3) Series F and G bonds were replaced by similar bonds, J and K respectively, yielding higher interest rates and eligible for annual purchases up to $200,000 maturity value for the two series combined for each purchaser When Series E bonds began to mature on May 1, 1951, owners of the matured bonds were offered an extension of maturity of 10 years or an exchange into Series G bonds as alternatives to cash redemption For Series E bonds maturing after April 30, 1952, a 10-year extension of maturity at an effective interest rate of approximately 3 per cent or an exchange into Series K bonds became available.

as against $3.8 billion of Series F bonds. Large-scale acquisition of F and G bonds by financial institutions and other corporations was prevented by the limitation on each subscriber's annual purchases.

The accrued discount on all unredeemed United States savings bonds held by individuals amounted to only $91 million in the aggregate in the first six years of the program, 1935–1940. Because sales of the bonds expanded after American entrance into World War II and because the effective interest rate on each bond rose for every additional six-month period that it was held (after the first year), the annual accruals of interest (discount) increased substantially thereafter reaching $605 million in 1950 and $697 million in 1952 (Table 14). In the years 1945–1950 accruals on savings bonds owned by individuals were equal to between 9.7 and 10.5 per cent of the annual totals of monetary interest received by individuals and to between 35 and 43 per cent of the annual totals of interest from all sources actually reported on taxable individual returns.

TABLE 14

ACCRUED DISCOUNT OF INDIVIDUALS ON UNREDEEMED UNITED STATES SAVINGS BONDS, BY CALENDAR YEARS, 1935–1952

(Millions of Dollars)

Year	Amount
1935	—
1936	3
1937	9
1938	16
1939	25
1940	38
1941	57
1942	78
1943	141
1944	253
1945	377
1946	440
1947	490
1948	554
1949	591
1950	605
1951	682
1952	697

Source: *Treasury Bulletin* and letter from Treasury Department. It was assumed that individuals held four-fifths of the amount of Series A–D bonds, all of Series E, and two-thirds of Series F.

Nevertheless, the lawful postponement of income tax accounting for the accruing interest (discount) on unredeemed United States savings bonds offers only a partial explanation for the discrepancy between total personal interest and the amounts reported on taxable returns. The most extreme assumption that can be made in this direction is that all the unredeemed savings bonds were owned by persons filing taxable returns and that the latter reported none of the current accruals of interest. If we add these accruals to the amounts of interest included on taxable returns, the proportion of total personal interest income that can be accounted for is raised somewhat, but only to a level averaging slightly more than 35 per cent in the 1940's as against much larger proportions in the 1920's, when the higher personal exemptions and credits excluded the vast majority of adults from income tax liability.

Our figures for the proportion of personal monetary or explicit interest reported on taxable returns for the years prior to 1929 are only roughly comparable with those for subsequent years because the two series are derived from the work of different investigators who employed different procedures. The two series overlap for the year 1929, when the proportion based on the Kuznets data was 41.8 per cent and that based on the Department of Commerce figures was 38.8 per cent. In 1950 the unadjusted proportion based on the Department of Commerce figures was only 25.9 per cent; and when adjusted to include all accrued interest on unredeemed savings bonds, it was 35.6 per cent. But in 1929 the personal exemptions were such that a married man without dependent children paid no income tax unless his net income was above $3,500, whereas in 1950 he was taxable if his adjusted gross income was as little as $1,333. Clearly the proportion of total interest income reported on taxable returns in recent years was much smaller than might reasonably have been expected.

5 *Interest from tax-exempt securities*

To determine how much of the discrepancy between total personal interest receipts and the amounts reported on taxable returns was attributable to interest received on tax-exempt securities, it was necessary to estimate individual receipts of such tax-exempt interest (Table 15).

After rising from $85 million in 1913 to $651 million in 1921, the estimated amount of wholly tax-exempt interest received by individuals fluctuated between $565 and $636 million during the next eleven years. Thereafter it declined without interruption to $239 million in 1947, after which it rose slightly and amounted to $274 million in 1952.

TABLE 15

WHOLLY TAX-EXEMPT INTEREST RECEIVED BY INDIVIDUALS, 1913–1952

(Millions of Dollars)

Year	State and local govt. securities	Wholly exempt			Direct and guar. fed. securities nominally only partially exempt	Total
		Direct federal securities	Postal savings deposits	Federal farm loan securities		
	(1)	(2)	(3)	(4)	(5)	(6)
1913	65.8	18.3	0.6			84.7
1914	75.2	18.3	0.8			94.3
1915	86.9	18.3	1.1			106.3
1916	99.2	18.0	1.6			118.8
1917	111.9	46.0	2.0			159.9
1918	125.1	70.6	2.3	2.4	86.0	286.4
1919	147.6	70.9	2.3	4.8	330.2	555.8
1920	175.2	80.0	2.2	7.3	373.0	637.7
1921	205.9	74.5	2.2	14.6	353.5	650.7
1922	235.4	60.5	2.1	26.2	299.3	623.5
1923	260.2	56.2	2.1	35.6	280.8	634.9
1924	289.6	56.2	2.2	42.6	207.5	598.1
1925	312.4	52.7	2.3	49.5	193.5	610.4
1926	332.9	52.1	2.3	54.0	154.1	595.4
1927	355.3	52.1	2.5	58.4	114.2	582.4
1928	377.6	52.1	2.6	60.5	86.5	579.3
1929	389.8	52.1	2.6	60.4	60.0	564.9
1929	389.8	52.1	2.8	60.4	60.0	565.1
1930	427.7	57.7	3.5	60.3	42.4	591.6
1931	454.3	56.1	7.2	58.0	36.2	611.8
1932	469.6	60.0	13.6	53.2	39.1	635.5
1933	439.8	62.7	19.2	52.8	35.8	610.3
1934	404.6	62.2	21.7	45.7	49.1	583.3
1935	374.5	34.1	21.9	39.2	60.5	530.2
1936	363.8	17.5	22.5	35.6	62.5	501.9
1937	346.1	18.8	23.1	29.9	70.6	488.5
1938	329.3	16.4	23.2	28.1	63.9	460.9
1939	308.3	14.7	23.3	28.2	66.4	440.9
1940	297.1	13.8	23.5	26.4	58.6	419.4
1941	273.5	12.3	3.9	22.9	51.5	364.1
1942	266.2	11.6		20.9	55.0	353.7
1943	261.0	11.9		20.1	55.6	348.6
1944	260.2	10.6		14.1	48.1	333.0
1945	256.1	6.6		4.5	35.2	302.4
1946	235.6	3.8			21.3	260.7
1947	217.0	3.5			18.6	239.1
1948	229.3	3.5			19.8	252.6
1949	232.3	3.4			19.5	255.2
1950	236.3	3.4			13.6	253.3
1[illegible]1	240.8	3.2			11.4	255.4

Estimated as follows

Col. 1: To George E Lent's estimates of individuals' holdings of state and local government securities, as presented in his *The Ownership of Tax-Exempt Securities, 1913–1953* (National Bureau of Economic Research, Occasional Paper 47, 1955, Appendix C), averaged for each pair of June 30 figures to obtain calendar year estimates, we applied interest rates obtained from estimates of the Dept of Commerce for 1913, 1922, and 1929–1950. Interest rates for the fiscal years between 1913 and 1922, and between 1922 and 1929 were estimated by straight-line interpolation and then averaged to obtain calendar year estimates For 1929–1950, June 30 debt estimates in the *Survey of Current Business*, October 1950, September 1952, were averaged to obtain calendar year figures, and the interest rates were derived by using the national income estimates of state and local government interest paid.

Col 2 As a result of discussions with Treasury officials and others, we estimated that individuals received 80 per cent of the interest from the wholly tax-exempt Treasury bonds outstanding in each year and 80 per cent of the interest from 3¾ per cent Victory Liberty Loan Notes of 1919–1922; and that 10 per cent of individuals' total holdings of wholly tax-exempt obligations of the federal government, as estimated by Lent (op cit), consisted of other Treasury notes and Certificates of Indebtedness The interest on wholly tax-exempt federal bonds was computed from the list of such bonds in the *Annual Report of the Secretary of the Treasury* each year, with adjustments for a part of a year when a security was issued or redeemed during the year For Treasury notes (other than the Victory Liberty 3¾'s cited above) and Certificates of Indebtedness, also listed in the *Annual Report*, an average rate of interest for the calendar year was computed and applied to calendar year averages of individuals' June 30 estimated holdings.

Col. 3 Figures for deposits in 1913–1918 were obtained for fiscal years from the *Annual Report of the Comptroller of the Currency* and averaged to obtain calendar year estimates Interest paid in these years was estimated by assuming the same rate between deposits and interest as obtained in calendar year 1920. Interest figures for 1919–1929 were taken from Simon Kuznets, *National Income and Its Composition, 1919–1938*, and for 1929–1941 from interest credited to depositors as reported in the *Annual Report of the Postmaster General* and averaged to obtain calendar year estimates. Tax exemption was withdrawn from postal savings interest on balances deposited after February 1941, and tax-exempt interest in the latter year was estimated at one-sixth of the total interest for the year

Col 4· From wholly tax-exempt Federal Farm Loan Securities. An average interest rate was derived for each calendar year from Federal Land Bank and Joint Stock Land Bank bonds outstanding and listed in the *Annual Report of the Farm Credit Administration* and the *Annual Report of the Federal Farm Loan Board* The interest rates were applied to calendar year averages of Lent's June 30 estimates of Federal Farm Loan holdings by individuals (Lent, *op cit*) Interest rates for 1923–1927 and 1943 were estimated by straight-line interpolation and 1944 and 1945 fiscal years were averaged to obtain a calendar year rate

Col 5 From nominally only partially tax-exempt direct and guaranteed federal securities Partially tax-exempt direct and guaranteed federal securities consist of bonds issued between 1917 and 1941 by the Treasury and several government corporations, the interest on the first $5,000 principal amount of which in the hands of any investor is exempt from both normal and surtaxes, and the interest from any additional amounts of which is exempt only from normal tax.

Interest rates were obtained by calendar years from the list of such securities in the *Annual Report of the Secretary of the Treasury* and were applied to calendar-year averages of Lent's June 30 estimates of individual holdings of partially tax-exempt federal direct and guaranteed securities (Lent, *op. cit*) less individual holdings of partially tax-exempt United States savings bonds (see note to Table 14)

For the years 1924–1927 and 1932–1943, both inclusive, we estimated that the amount of partially tax-exempt interest enjoying full tax exemption constituted the same proportion of the total as was tabulated and reported in *Statistics of Income* for taxpayers with net incomes of $5,000, $6,000, or $10,000 and over in different years. Similar figures were not published for other years. We estimated the proportion for the years 1918–1923 by adding several percentage points to the 1924 proportion to allow for the fact that for several years after World War I complete exemption from income tax was extended for limited periods to the interest on varying principal amounts in excess of $5,000 if the holder subscribed to additional amounts of certain new Treasury issues. We estimated the proportions for the years 1928–1931 by straight-line interpolation between the figures for 1927 and 1932 and we assumed a level proportion for the years 1944–1950 slightly below that for 1943.

For the five years 1918–1923, Treasury obligations were the most important source of tax-exempt interest for individuals. Although the interest on most of the federal securities issued during World War I was nominally exempt only from normal tax, except that interest on the first $5,000 principal amount held by any taxpayer was exempt also from surtaxes, complete tax exemption was granted for limited periods on substantial principal amounts of particular issues to promote subscriptions. The Revenue Act of 1921 consolidated the various previous temporary tax exemptions of the 4 per cent and 4¼ per cent Liberty bonds into an exemption from normal and surtax of the interest from an aggregate principal amount of $125,000 for each investor for two years after the termination of the war (July 2, 1921) and from a principal amount of $50,000 for three additional years. Such temporary exemptions were responsible for most of the tax-exempt interest derived by individuals from federal securities in 1919–1926. The effect of the reduction, and then the expiration, of the temporary grants of complete exemption was intensified by the substantial retirements of federal securities that took place during the twenties. By 1931, wholly tax-exempt interest obtained by individuals from direct and guaranteed federal obligations, including Postal Savings deposits, had shrunk to $100 million as compared with a high of $455 million in 1920.

The decline in wholly tax-exempt interest of individuals from federal obligations during the 1920's was partly offset by a steady rise in that from state and local government securities. Individual receipts of such interest rose each year without interruption from $66 million in 1913 to $470 million in 1932. In the following fifteen years, however, the annual totals fell steadily, largely because of debt retirements and declining interest rates, and reached a low of $217 million in 1947.

The long period of deficit financing by the federal government in the 1930's and 1940's was accompanied, after 1935, by a decline rather than an increase in wholly tax-exempt interest income of individuals from federal securities. The reasons were that the fully exempt securities issued by the Treasury consisted of short-term obligations commanding low interest rates and these securities were mainly purchased by institutional investors; some previous issues of fully exempt securities were retired; and the federal government ceased issuing tax-exempt securities in 1941.

Since the total of individuals' receipts of tax-exempt interest fell by more than one-half in 1932–1950 while their total monetary interest income rose about 34 per cent, it seems clear that the changes in their

receipts of tax-exempt interest should have narrowed, not widened, the gap between their total receipts of monetary interest and the amount reported on taxable returns.

If we subtract from the annual totals of monetary interest of individuals the sum of their tax-exempt interest and accrued interest on unredeemed United States savings bonds and compare the remainders with the annual amounts of interest reported on both taxable and nontaxable individual returns, we obtain the percentages in column 6 of Table 16; and if the comparison is with the amount of interest reported *only* on taxable returns, we obtain the percentages in the last column of the table. After these adjustments the proportion of personal interest income reported on individual tax returns in the 1940's remains much lower than in the 1920's, and only moderately higher than in the late 1930's.

6. *Underreporting of interest income*

Some part of the recent wide gap between the amount of interest that individuals are estimated to receive and the amount they report on taxable returns is doubtless due to receipts by children, retired persons, and others whose total net incomes are too small to be taxable. Another part can be attributed to nonprofit institutions, whose interest income in the aggregate has been growing[18] and whose interest receipts are treated as those of individuals in the national income accounts published by the Department of Commerce. Excluded from taxable interest each year is the amount received by individuals whose ordinary income has been offset by deductible capital and business losses. There may also be minor undiscovered statistical offsets. On the other hand, the gap is understated in this respect: the figures for total monetary interest receipts of individuals are exclusive of interest received by some individuals from other individuals. But on the whole the figures strongly suggest that a significant fraction of the unreported interest estimated to have been received by individuals represents tax evasion.

Many taxpayers among the millions added to the income tax rolls since World War II doubtless neglected to report interest received on savings bank deposits, redeemed savings bonds, building and loan association shares, and mortagage and other loans. The sample study of tax returns for 1948, conducted by the Bureau of Internal Revenue

[18] Robert Rude, in a preliminary progress report of his study of private nonprofit institutions in the United States, found that the endowments (other than physical facilities) of private educational institutions and private nonprofit hospitals had more than doubled between 1938 and 1947, reaching $2.8 billion in the latter year (National Bureau of Economic Research unpublished staff progress reports, May 1952).

TABLE 16

TOTAL MONETARY INTEREST RECEIPTS, TAX-EXEMPT INTEREST, ACCRUED INTEREST ON UNITED STATES SAVINGS BONDS, AND INTEREST REPORTED ON TAXABLE RETURNS, 1918–1950

	Total monetary interest	Wholly tax-exempt interest received by individuals	Accrued interest of individuals on unredeemed U.S. savings bonds	Interest not tax-exempt or on unredeemed savings bonds	Interest reported on taxable and nontaxable returns	All reported interest as percentage of col (4) (per cent)	Interest reported on taxable returns	Taxable reported interest as percentage of col. (4) (per cent)
	(Figures other than percentages are for millions of dollars)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1918	1,967	286		1,681	1,403	83.5	1,306	77.7
1919	2,585	556		2,029	1,501	74.0	1,395	68.8
1919	2,932	556		2,376	1,501	63.2	1,395	58.7
1920	3,333	638		2,695	1,642	60.9	1,482	55.0
1921	3,522	651		2,871	1,690	58.9	1,262	44.0
1922	3,589	624		2,965	1,858	62.7	1,442	48.6
1923	3,772	635		3,137	2,330	74.3	1,869	59.6
1924	3,900	598		3,302	2,414	73.1	2,009	60.8
1925	4,055	610		3,445	1,941	56.3	1,641	47.6
1926	4,120	595		3,525	2,083	59.1	1,757	49.8
1927	4,299	582		3,717	2,213	59.5	1,834	49.3
1928	4,569	579		3,990	2,398	60.1	2,032	50.9
1929	4,831	565		4,266	2,534	59.4	2,018	47.3
1929	5,198	565		4,633	2,534	54.7	2,018	43.6
1930	4,795	592		4,203	2,201	52.4	1,708	40.6
1931	4,843	612		4,231	1,703	40.3	1,141	27.0
1932	4,645	636		4,009	1,439	35.9	1,012	25.3
1933	4,421	610		3,811	1,224	32.1	848	22.3
1934	4,239	583		3,656	1,114	30.5	814	22.3
1935	4,039	530		3,509	1,114	31.7	813	23.2
1936	3,886	502	3	3,381	1,061	31.4	903	26.7
1937	3,883	488	9	3,386	1,035	30.6	882	26.0
1938	3,822	461	16	3,345	1,055	31.5	866	25.9
1939	3,761	441	25	3,295	1,072	32.5	900	27.3
1940	3,701	419	38	3,244	1,206	37.2	988	30.5
1941	3,672	364	57	3,251	1,306	40.2	1,148	35.3
1942	3,494	354	78	3,062	1,264	41.3	1,142	37.3
1943	3,391	349	141	2,901	1,132	39.0	1,095	37.7
1944	3,451	333	253	2,865	1,124	39.2	1,080	37.7
1945	3,903	302	377	3,224	1,109	34.4	1,069	33.2
1946	4,376	261	440	3,675	1,289	35.1	1,228	33.4
1947	4,874	239	490	4,145	1,341	32.4	1,246	30.1
1948	5,256	253	554	4,449	1,492	33.5	1,301	29.2
1949	5,737	256	591	4,891	1,745	35.7	1,515	31.0
1950	6,243	253	605	5,385	1,825	33.9	1,619	30.1

Source: See notes to Tables 1, 14, and 15.

Col. 1: See notes to Table 1.

Col. 2: See notes to Table 15.

Col. 3: See notes to Table 14.

under its Audit Control Program, indicated that if all the returns of that year had been audited, about 2 million of them would have been found to contain errors in interest and that in about 1 million of them interest would have been found to be the major source of error[19] Unfortunately, the study did not yield measures of the amount of understatement of interest or other components of income. The frequency of interest errors was about four times that of errors in dividends. In 93 per cent of the returns where interest was the major source of error, there was understatement of tax liability. Although returns with incomes under $7,000 accounted for about four-fifths of those showing interest errors, such errors were not unimportant among returns with larger incomes More than half of the estimated increase in tax liability estimated to be disclosable on returns in which interest was the major source of error was attributed to individuals reporting adjusted gross incomes of $7,000 or more or gross receipts from business or profession of $25,000 or more.

The audit sample therefore indicates that understatement and unreporting of personal interest income is not confined to the failure of thousands of small depositors and other small interest recipients to account fully for trifling individual sums. Also suggesting that sizable amounts of interest received from mortgages and federal and corporate obligations have been unreported for tax purposes is the fact that these sources appear to account for not far from three-quarters of the total direct monetary interest paid to individuals in recent years. Selma Goldsmith in *Volume Thirteen* of the National Bureau's *Studies in Income and Wealth* ([1951], pp. 303–304) presented rough estimates, reproduced in Table 17 of the sources of direct interest income of individuals in the three years 1944–1946. Excluding tax-exempt interest, accrued interest on unredeemed United States savings bonds, and residuals varying in the different years between 9 and 18 per cent

Col 4 Equals col 1 minus the sum of the figures in cols 2 and 3

Col. 5 See notes to Table 4 For years 1918 through 1930, includes annuities and dividends from foreign corporations other than those obtaining more than 50 per cent of their gross income from sources in the United States For 1931, interest on returns with no net income was estimated from ratio obtaining for taxable returns between interest and other income For 1941 through 1943, when interest was lumped together on 1040a returns with dividends rents, royalties, and annuities, interest income on these returns was estimated by assuming that it constituted the same proportion of the lumped components as on 1040 returns The interest figures for 1943 include partially tax-exempt dividends. For 1944 and 1945, when dividends and interest were lumped together in the returns, interest was estimated to constitute the same proportion of the total of dividends and interest as in 1946

Col 7 See notes to Table 4

[19] See Marius Farioletti, "Some Results of the First Year's Audit Control Program of the Bureau of Internal Revenue," *National Tax Journal*, March 1952, also *The Audit Control Program* (Bureau of Internal Revenue, 1951.

TABLE 17

SOURCES OF DIRECT MONETARY INTEREST INCOME OF INDIVIDUALS,[a] 1944–1946

(Millions of Dollars)

Source	1944	1945	1946
Mortgage loans	650	670	780
Federal obligations[b]	285	575	1,025
Corporate bonds	730	700	650
Commercial bank time deposits	200	300	300
Mutual savings bank deposits	200	215	240
Savings and loan associations and credit unions	165	175	200
Unallocated residual	493	450	331
Total[c]	2,723	3,085	3,526

[a] Individuals include fiduciaries and nonprofit organizations furnishing services to individuals.

[b] Including postal savings deposits.

[c] Direct monetary interest income of individuals was derived by deducting imputed interest, tax-exempt interest, and accrued interest on unredeemed United States savings bonds from the interest component of personal income in the national income statistics of the Department of Commerce. These totals vary slightly from our figures used elsewhere in this book primarily because of small differences between Mrs. Goldsmith's estimates and ours of tax-exempt interest.

Source: Selma F. Goldsmith, "Appraisal of Basic Data for Size Distributions," *Studies in Income and Wealth, Volume Thirteen* (National Bureau of Economic Research, 1951), p. 304.

that she could not allocate to any of the six principal sources of interest income, she found that only about one-fourth of the remainder of monetary interest paid to individuals represented interest on bank deposits and interest or dividends paid by savings and loan associations in the three years. Since, for the three years as a whole, only about one-third of the total cash interest receipts of individuals, exclusive of tax-exempt interest and accrued interest on unredeemed savings bonds, was reported on taxable returns, it would seem that significant amounts of interest on mortgages and corporate and govenment obligations were not reported.

The figures suggest that greater underreporting of interest income developed after the sharp rise in tax rates and taxable incomes during World War II. In 1941 the amount of interest on taxable returns was 35 per cent of total monetary interest received by individuals, exclusive of tax-exempt interest and accrued interest on unredeemed United States savings bonds. This percentage fell by more than 6 percentage points to slightly above 29 by 1948 and was a little above 30 in 1950 (Table 16).

When the Treasury Department in September 1952 proposed to require banks to file information returns with respect to all payments of interest amounting to $100 or more in one year, instead of requiring such returns only with respect to interest payments of $600 or more, the *Wall Street Journal* (Feb. 18, 1953) reported that some holders of savings accounts were dividing up their accounts among a number of banks in order that their interest receipts from any one bank would fall short of $100 a year In December 1952 the Treasury withdrew the proposed change in its regulations because of protests from bankers that the change would greatly increase their paper work.

III. TAX REVENUES FROM INTEREST

1. *Tax revenues from interest income in the 1940's far exceeded those of previous decades, with a growing proportion from the lower taxable incomes*

Since the personal income tax is levied at the same rates on the total from all sources (except for tax-exempt interest and long-term capital gains) of an individual's taxable income, less his personal exemptions and allowable deductions, his tax liability is not strictly made up of separate amounts attributable to each of the various components of his income, such as salaries, dividends, rents, etc.[20] Nevertheless it is useful for some purposes to act as if this were the situation. In practice the additional tax that would be payable on an additional increment of income is often taken into account by investors in choosing the character and form of new investments—as between bonds and common stocks and as between tax-exempt and taxable bonds, for example. This aspect receives attention in the next section of this paper. Because the composition of taxable incomes of the different income groups and the structure of tax rates have varied significantly over a period of years, it is also interesting to measure the varying average impact of the income tax upon the different components of personal incomes. The average effective tax rate for each income group can be calculated by dividing its total tax liability by its total adjusted gross income. By

[20] The effective tax rates on some kinds of income have differed from the formal rates in varying degree from time to time because of exemptions or credits applicable to particular kinds of income. A zero rate, in effect, has been applied without interruption to interest from state and local government obligations and from some obligations of the federal government and its instrumentalities. Interest from large amounts of securities issued by the federal government and its agencies prior to 1941 was exempt from normal tax, although not from surtax An earned income credit, varying in different years between 1924 and 1943 (not allowable in 1932 and 1933), reduced the effective tax rates on incomes received as compensation for personal services and with different limits, on property income also. Prior to 1936, dividend income was exempt from normal tax Under the Revenue Act of 1954, the first $50 of dividend income received by any taxpayer was made wholly exempt from income tax, and a tax credit of 4 per cent of the remainder of his dividend income was made applicable against his total tax liability.

applying the average effective tax rate so obtained to the amount of interest income reported by each income group, a measure of the tax liability that can be attributed to the interest component of taxable incomes may be obtained.[21] In this way interest is not treated as the marginal component of income, subject to the highest applicable bracket rates, but treated as if it were distributed among the various brackets in the same proportions as all other components of income and subject only to the average effective tax rate paid by each person or income group on his or its adjusted gross income as a whole.

Measured in this way, the amount of tax revenue supplied by the interest component of taxable individual incomes in 1918–1950 ranged from a low of $35 million in 1931 to a high of $288 million in 1950 (Table 18). In 1918–1920 low personal exemptions, high tax rates, and exceptionally high interest rates (the average yield of Moody's Aaa bonds was 5.49 per cent in 1919 and 6.12 per cent in 1920) contributed to the relatively large revenues from interest in 1918–1920 ($147, $135, $104 million). During 1921–1929 the net effect of more liberal personal exemptions, falling tax rates, and sustained or rising interest income and total income was to produce a fairly stable plateau of income tax revenue from interest, averaging $79 million annually. The influence of liberalized personal exemptions in 1921 and 1925 is clearly reflected in the declines of those years.

During the 1930's, largely because of the depressed levels of total personal incomes, tax revenues from interest averaged about 36 per cent lower than during the twenties, despite the reduction in personal exemptions and increased in tax rates that began in 1932. A sharp increase in income tax revenues from interest took place during the 1940's with the steep reductions in personal exemptions, increases in tax rates, and growth in the public debt that acccompanied the rearmament program and World War II. In 1950 the $288 million of income tax liability for interest receipts reported on taxable individual returns was more than six times that of 1938 and greater than that of any previous year.

Reflecting the progressive structure of effective income tax rates, inclusive of the personal exemption provisions, the great bulk of the income taxes attributable to interest receipts in all years between 1918 and 1940 was levied on the upper income groups (Table 19). Taxpayers with incomes of $10,000 or more in this period, though receiving about 46 per cent on the average of the annual amounts of interest

[21] For some purposes the marginal or highest applicable bracket rate is more significant than the average effective tax rate, as is illustrated presently.

TABLE 18

INTEREST TAX LIABILITY, BY INCOME GROUPS, 1918–1950

(Thousands of Dollars)

	Net income classes[a]									
	Less than 2	2–3	3–5	5–10	10–25	25–50	50–100	100–590	500 and over	Total
1918	1,706	2,261	5,675	9,295	17,500	17,322	20,367	45,649	27,647	147,242
1919	1,125	1,168	3 990	7,104	16,626	17,410	20,745	45,849	21,290	135,307
1920	1,399	1,995	4,475	7,707	17,031	16,682	18,281	26,090	10,788	104,452
1921	1,313	1,532	3,166	6,536	14,934	14,044	13,620	16,595	4,878	76,618
1922	1,204	1,559	3,373	7,132	15,697	16,473	18,108	24,375	8,339	96,258
1923	1,434	1,631	4,221	5,562	12,548	12,761	12,688	16,372	5,614	72,831
1924	807	1,009	2,646	2,595	9,382	13,149	15,381	21,908	7,778	74,655
1925	134	365	804	1,705	8,388	12,984	14,066	17,422	6,886	62,755
1926	135	417	756	1,961	9,182	13,666	15,029	19,546	7,600	68,294
1927	89	376	662	2,190	9,897	14,900	16,745	22,890	8,988	76,737
1928	141	490	801	2,313	10,402	15,922	19,076	28,715	16,383	94,242
1929	44	148	254	970	7,869	14,364	17,131	27,341	18,124	86,245
1930	104	415	595	2,148	8,951	12,598	13,085	16,267	7,147	61,310
1931	75	299	412	1,615	6,456	7,542	7,712	7,635	3,329	35,074
1932	992	1,057	1,837	4,821	10,186	11,469	9,789	9,577	2,336	52,068
1933	909	874	1,557	4,059	8,485	9,271	8,204	9,175	2,582	45,118
1934	534	694	1,227	4,376	10,889	10,429	8 401	7,994	2,725	47,266
1935	497	659	1,165	4,164	11,191	11,167	9,842	9,603	2,448	50,736
1936	796	991	1,667	4,110	10,896	12,031	12,730	15,913	6,421	65,556
1937	774	900	1,645	4,131	10,759	11,042	11,214	14,657	6,498	61,620
1938	987	943	1,647	4,060	9,607	8,682	7,110	9,123	4,447	46,606
1939	1,017	1,039	1,776	4,115	10,093	9,966	8,516	11,758	6,040	54,320
1940	1,823	1,294	2,366	5,022	13,661	15,291	12,261	13,563	6,065	71,345
1941	5,871	5,185	9,959	13,954	28,028	23,424	16,941	17,462	5,386	126,211
1942	16,011	11,607	17,937	24,483	42,909	33,922	23,296	21,400	4,544	196,109
1943	16,101	12,549	20,014	31,746	51,050	39,915	29,129	26,835	5,317	232,657
1944	11,422	11,462	20,678	33,553	55,344	42,588	33,042	29,232	10,168	247,489
1945	10,013	11,218	18,488	34,285	58,691	47,042	33,855	28,550	9,694	251,837
1946	9,442	10,250	17,748	32,414	61,081	49,862	38,284	34,909	11,297	265,288
1947	9,918	11,238	18,102	33,541	60,322	49,176	36,603	35,802	11,167	265,867
1948	4,187	6,625	14,970	28,106	45,612	40,519	34,847	38,739	11,827	225,432
1949	5,884	8,392	18,891	33,618	49,050	41,747	34,921	36,099	11,519	240,122
1950	5,445	7,828	18,520	36,819	58,210	54,271	44,351	47,083	15,528	288,054

[a] In 1944 changed to Adjusted Gross Income Classes.

Source *Statistics of Income* To the estimated tax liability arrived at as described in the text, we added the tax liability attributable to the interest component of income from fiduciaries received by individuals and beneficiaries This added tax liability was estimated by the same procedures as those described in the notes to Table 4 However, a slight overestimate of tax liability attributable to interest, of the order of 1 per cent results from the small duplication in the figures

TABLE 19

SHARES OF INCOME GROUPS IN TOTAL TAX LIABILITY ATTRIBUTABLE TO INTEREST, 1918–1950

(Percentages)

	Net income classes* (thousands of dollars)								
	Less than 2	2–3	3–5	5–10	10–25	25–50	50–100	100–500	500 and Over
1918	1.2	1.5	3.9	6.3	11.9	11 8	13 8	31.0	18.7
1919	0.8	0.9	2.9	5 3	12 3	12.9	15.3	33.9	15.7
1920	1.3	1.9	4.3	7.4	16.3	16.0	17.5	25.0	10.3
1921	1.7	2.0	4.1	8.5	19.5	18.3	17.8	21.7	6.4
1922	1.3	1.6	3.5	7.4	16 3	17.1	18.8	25.3	8.7
1923	2.0	2.2	5.8	7.6	17.2	17.5	17.4	22 5	7.7
1924	1.1	1.4	3.5	3.5	12.6	17.6	20 6	29.3	10.4
1925	0.2	0.6	1.3	2.7	13.4	20.7	22.4	27.8	11.0
1926	0.2	0.6	1.1	2.9	13.4	20 0	22.0	28.6	11 1
1927	0.1	0.5	0.9	2.9	12 9	19.4	21 8	29.8	11.7
1928	0.1	0.5	0.8	2 5	11.0	16.9	20.2	30 5	17.4
1929	0.1	0.2	0.3	1.1	9 1	16.7	19.9	31.7	21.0
1930	0 2	0.7	1.0	3.5	14 6	20.5	21.3	26.5	11.7
1931	0.2	0.9	1.2	4 6	18 4	21.5	22.0	21 8	9.5
1932	1.9	2.0	3.5	9.3	19.6	22 0	18 8	18.4	4.5
1933	2 0	1.9	3.5	9.0	18.8	20.5	18.2	20.3	5.7
1934	1.1	1.5	2.6	9.3	23.0	22.1	17.8	16 9	5.8
1935	1.0	1.3	2.3	8 2	22.1	22.0	19.4	18.9	4.8
1936	1.2	1.5	2.5	6 3	16 6	18.4	19 4	24.3	9 8
1937	1 3	1.5	2.7	6.7	17.5	17.9	18.2	23.8	10.5
1938	2.1	2 0	3.5	8.7	20.6	18 6	15 3	19.6	9 5
1939	1.9	1 9	3.3	7.6	18.6	18.3	15.7	21.6	11.1
1940	2 6	1.8	3.3	7.0	19.1	21.4	17.2	19.0	8 5
1941	4.7	4.1	7.9	11.1	22.2	18.6	13.4	13.8	4.3
1942	8.2	5.9	9.1	12.5	21.9	17.3	11.9	10 9	2 3
1943	6 9	5.4	8.6	13.6	21.9	17.2	12 5	11.5	2.3
1944	4.6	4.6	8.4	13.6	22.4	17.2	13.4	11.8	4.1
1945	4.0	4.5	7.3	13.6	23.3	18.7	13.4	11.3	3 8
1946	3.6	3.9	6.7	12.2	23 0	18.8	14.4	13.2	4 3
1947	3.7	4.2	6.8	12.6	22.7	18.5	13.8	13.5	4.2
1948	1.9	2.9	6.6	12.5	20.2	18.0	15.5	17.2	5 2
1949	2.5	3.5	7.9	14.0	20.4	17.4	14.5	15.0	4 8
1950	1.9	2.7	6.4	12.8	20.2	18.8	15.4	16.3	5.4

* In 1944 changed to Adjusted Gross Income Classes.

Note: The figures do not necessarily add to 100 per cent due to rounding.

reported on taxable returns, paid on the average 89 per cent of the annual amounts of income taxes attributable to interest. This extreme relationship was moderated after 1940, partly because of a change in the distribution of reported interest income in favor of the lower income groups and partly because of the relatively greater increases of effective tax rates on smaller than larger incomes. Between 1941 and 1949 inclusive, taxpayers with incomes of $10,000 or more reported an annual average of 41 per cent of all the interest on taxable returns and paid an annual average of 71 per cent of the income taxes attributable to interest.

2. *Taxes on interest income declined radically as a share of total income taxes, but absorbed one-sixth or more of all interest reported on taxable returns in most years since 1942.*

As a proportion of total personal income tax liability, the amount contributed by interest income ranged during the twenties from 11.2 to 8 1 per cent (Table 20) The proportion rose to 15 8 per cent between 1929 and 1932, when other components of income shrank more drastically during the economic collapse of that period. Thereafter a sharp fall took place that continued with unimportant interruptions for the next twelve years. In 1950, income tax liability attributable to interest accounted for only 1 6 per cent of total personal income tax liability.

Income taxes absorbed a relatively small proportion of total interest income reported on taxable returns during the two decades 1921–1940, the proportion varying between 3 1 and 7.3 per cent (Table 21). It rose sharply in the years following 1940, climbing to 11 0 per cent in 1941, 17.2 in 1942, 21 2 in 1943, 22 9 in 1944, and 23.6 in 1945, after which it gradually declined to 15 9 per cent in 1949 and then rose to 17.8 per cent in 1950

IV. INTEREST INCOME AND INTEREST RATES AFTER TAXES

For the upper income groups, of course, the proportion of interest income absorbed by income taxes rose to much higher levels during the 1940's, and the marginal rates rose higher still. The effective federal tax rates on the first dollar of additional interest or other ordinary income for a married man with two dependent children and the cited amounts of statutory net income from other sources for the years 1940–1954 are shown in Table 22. To these, in states imposing their own income taxes, should be added the marginal tax rates under the state laws.

TABLE 20

TAX LIABILITY ATTRIBUTABLE TO INTEREST AS A PERCENTAGE OF TOTAL TAX LIABILITY, BY INCOME GROUPS, 1918–1950

(Percentages)

	Net income classes* (thousands of dollars)									
	Less than 2	2–3	3–5	5–10	10–25	25–50	50–100	100–500	500 and Over	Average
1918	6.4	6.4	6.8	10.0	12.3	13.3	13.8	14.7	17.3	13.1
1919	4.6	4.1	5.3	7.8	10.1	11.2	11.1	12.5	12.1	10.7
1920	3.8	4.4	5.4	7.9	9.9	10.8	11.2	11.5	11.4	9.7
1921	4.5	7.4	7.4	9.5	11.8	12.4	11.8	11.4	8.6	10.7
1922	4.4	7.5	7.1	10.1	12.7	13.1	12.6	11.4	9.5	11.2
1923	7.7	9.8	9.2	10.3	12.1	12.3	11.7	11.0	9.2	11.0
1924	7.6	9.9	9.8	9.0	12.0	12.0	11.3	10.2	8.7	10.6
1925	7.6	9.6	9.7	8.9	11.3	10.8	9.5	7.3	5.7	8.5
1926	7.4	9.9	10.4	9.7	12.7	12.1	10.7	8.2	5.6	9.3
1927	7.0	9.5	10.2	10.6	13.3	12.5	10.7	8.0	5.5	9.2
1928	8.8	11.4	10.7	10.1	12.6	11.7	9.8	7.0	5.4	8.1
1929	7.7	10.5	10.5	10.2	13.1	12.6	10.7	7.7	6.1	8.6
1930	8.0	12.5	11.1	12.3	18.1	17.3	15.0	11.3	7.5	12.9
1931	7.8	12.1	10.8	13.0	20.2	18.8	17.2	11.3	7.9	14.3
1932	8.0	10.8	8.8	13.5	20.3	26.3	20.8	12.8	6.6	15.8
1933	8.7	11.3	8.5	11.6	15.4	17.7	14.3	10.3	5.3	12.1
1934	6.1	9.2	6.7	10.2	13.0	12.3	9.9	6.8	4.3	9.2
1935	4.9	7.1	5.6	8.5	10.8	10.5	8.7	5.8	3.1	7.7
1936	5.6	7.1	5.2	5.2	6.2	6.3	5.9	4.8	4.1	5.4
1937	4.4	5.8	4.2	4.9	6.1	6.2	5.8	4.9	4.8	4.5
1938	6.2	6.8	4.9	5.7	7.2	7.2	6.1	5.4	4.9	6.1
1939	4.6	4.9	3.7	4.9	6.4	6.8	5.8	5.8	6.2	5.9
1940	3.0	2.6	3.0	4.2	5.4	5.6	4.0	4.6	5.2	4.8
1941	1.9	1.1	2.4	3.4	4.1	4.1	3.7	3.9	3.9	3.2
1942	1.3	0.7	1.3	2.7	3.5	3.6	3.1	3.1	2.7	2.2
1943	0.8	0.4	0.7	2.1	2.9	3.0	2.9	3.1	2.8	1.6
1944	0.6	0.4	0.5	1.7	2.7	3.0	3.2	3.5	5.4	1.5
1945	0.6	0.4	0.5	1.6	2.4	2.8	2.9	3.2	4.7	1.5
1946	0.7	0.4	0.5	1.5	2.3	2.7	3.0	3.6	4.3	1.6
1947	0.7	0.4	0.4	1.3	2.2	2.7	3.0	3.6	4.1	1.5
1948	0.6	0.4	0.4	0.9	2.0	2.6	2.7	3.1	3.5	1.4
1949	0.9	0.5	0.5	1.1	2.3	2.9	3.2	3.5	3.7	1.6
1950	0.8	0.5	0.4	0.9	2.3	2.8	2.9	2.9	3.0	1.6

* In 1944 changes to Adjusted Gross Income Classes.

TABLE 21

PROPORTION OF INTEREST INCOME ABSORBED BY INCOME TAX, TAXABLE RETURNS, 1918–1950

	Percentage of interest income absorbed by income tax		Percentage of interest income absorbed by income tax
1918	11 3	1935	6.2
1919	9.7	1936	7 3
1920	7 0	1937	7 0
1921	6 1	1938	5 4
1922	6 7	1939	6 0
1923	3 9	1940	7 2
1924	3 7	1941	11 0
1925	3 8	1942	17 2
1926	3 9	1943	21.2
1927	4 2	1944	22.9
1928	4 6	1945	23 6
1929	4 3	1946	21.6
1930	3 6	1947	21 3
1931	3.1	1948	17.3
1932	5.1	1949	15 9
1933	5 3	1950	17 8
1934	5 8		

1 *Did low interest rates and high tax rates deter fixed-interest investment?*

The pronounced increase in personal income tax rates in the 1940's took place after interest rates had declined markedly for most of a decade and were still falling In consequence the after-tax or "take-home" yields obtainable by individuals from fixed-interest investment suffered a sharp compound reduction. In 1929 a married man with two dependent children could obtain a marginal after-tax yield of 4 71 per cent from a moderate investment in the average of Moody's Aaa corporate bonds if his statutory net income from other sources was $5,000, a "take-home" yield of 4.21 per cent, if $25,000, and one of 3.93 per cent, if $50,000 By 1950 these after-tax yields had shrunk to 2.16, 1 71, and 1 21, respectively. Similar radical reductions in after-tax

TABLE 22

MARGINAL TAX RATES ON THE FIRST DOLLAR OF ADDITIONAL INTEREST OR OTHER ORDINARY INCOME, 1940–1954

(Married, with Two Dependent Children)

	Statutory net income (thousands of dollars)					
	$5	$10	$25	$50	$100	$1,000
1940[a]	4 4	11	34.1	48.4	66	78.4
1941	13	21	48	59	68	78
1942	22	34	58	69	83	88
1943[c]	24.8	36.8	60.8	71.8	88	90[d]
1944–1945	25	37	62	75	90	90[d]
1946–1947	20.9	32.3	56.1	68.4	82.7	86.5
1948–1949[b]	16 6	19 4	33.4	51.9	63.4	82.1
1950[b]	17.4	20	34 6	53.7	65.5	84.4
1951[b]	20 4	22 4	39	60	73	91
1952–1953[b]	22.2	24.6	42	66	75	92
1954[b]	20	22	38	59	72	91

[a] Includes Defense Tax.

[b] Rates are for joint return.

[c] Includes Victory Tax.

[d] Taking into account the statutory limitation of the maximum effective rate to 90 per cent.

For the years 1946–1954 varying maximum effective rates were imposed on the whole of net income that had the effect of reducing the marginal rates applicable to additional income in the uppermost income brackets. The levels of net income, for each year, at which the maximum effective rates became operative and the resulting effective rates were as follows.

Year	*Income*	*Per cent*
1946–1947	$2,700,000	85.5
1948–1949	1,013,379	77 0
1950	1,196,702	80.0
1951	1,357,685	87.2
1952–1953	1,219,400	88 0
1954	1,261,400	87 0[e]

[e] Assuming no dividend income.

yields occurred in bonds of lower quality. The shrinkage was even greater at higher levels of income (Tables 23 and 24, Charts 6 and 7).

Nevertheless, individual investors as a whole added very much larger amounts to their holdings of fixed-interest securities during the 1940's than in any previous decade. They expanded their ownership of federal securities from $10 1 billion at the end of 1939 to $66.3 billion at the end of 1950.[22] This increase was some $14 billion greater than the entire net federal debt outstanding at the end of 1939, about $12 billion greater than the entire amount of net long-term corporate debt outstanding on that date, and some $3 billion greater than the

[22] *Treasury Bulletin*, March 1956, p. 27

TABLE 23

MARGINAL AFTER-TAX YIELDS OF MOODY'S Aaa CORPORATE BONDS AT SELECTED NET INCOMES, 1919–1954
(Married, with Two Dependent Children)

Year	Moody's Aaa*	Selected net incomes				
		$5,000	$10,000	$25,000	$50,000	$100,000
1919	5.49	5.22	4.83	4.45	3.73	2.20
1920	6.12	5.81	5.39	4.96	4.16	2.45
1921	5.97	5 67	5.25	4.84	4.06	2.39
1922	5.10	4 90	4.59	4.18	3.52	2.24
1923	5.12	4.97	4.74	4.43	3.93	2.97
1924	5.00	4.90	4.75	4.35	3.80	2.85
1925	4.88	4 81	4.68	4.29	4.00	3.66
1926	4.73	4.66	4 54	4 16	3 88	3.55
1927	4.57	4.50	4.39	4.02	3.75	3.43
1928	4.55	4.48	4.37	4.00	3.73	3.41
1929	4.73	4.71	4.59	4.21	3 93	3.59
1930	4.55	4 48	4 37	4 00	3.73	3.41
1931	4.58	4 51	4 40	4.03	3.76	3.44
1932	5 01	4.81	4.51	4.11	3.46	2.20
1933	4 49	4.31	4.04	3.68	3.10	1.98
1934	4.00	3 84	3.64	3.24	2 76	1.84
1935	3.60	3.46	3 28	2 92	2 48	1.66
1936	3.24	3 11	2.95	2.62	2.24	1.33
1937	3 26	3.13	2.97	2.64	2.25	1.34
1938	3 19	3.06	2.90	2.58	2.20	1.31
1939	3 01	2 89	2.74	2.44	2.08	1.23
1940	2.84	2.72	2.53	1.87	1.47	0.97
1941	2 77	2.41	2 19	1 44	1.14	0.89
1942	2.83	2.21	1.87	1.19	0.88	0.48
1943	2 73	2 05	1.73	1.07	0 77	0.33
1944	2 72	2 04	1.71	1 03	0.68	0.27
1945	2 62	1.96	1.65	1.00	0.66	0.26
1946	2.53	2.00	1.71	1.11	0 80	0.44
1947	2.61	2 06	1.77	1.15	0.82	0 45
1948	2.82	2.35	2.27	1.88	1.36	1.03
1949	2.66	2 22	2.14	1.77	1.28	0.97
1950	2.62	2.16	2.10	1.71	1.21	0.90
1951	2.86	2 28	2.22	1.74	1.14	0.77
1952	2.96	2.30	2.23	1.72	1.01	0.74
1953	3.20	2.49	2.41	1.86	1.09	0.80
1954	2.90	2 32	2.26	1.80	1.19	0.81

* See Table 7.

TABLE 24

MARGINAL AFTER-TAX YIELDS OF MOODY'S Baa CORPORATE BONDS, AT SELECTED NET INCOMES, 1919–1954

(Married, with Two Dependent Children)

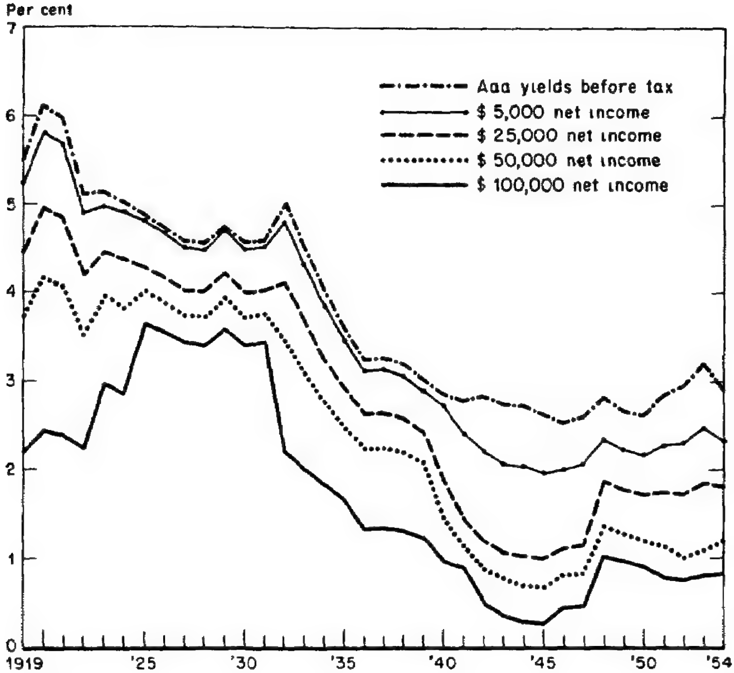
Year	Moody's Baa*	Selected net incomes				
		$5,000	$10,000	$25,000	$50,000	$100,000
1919	7.25	6.89	6.38	5.87	4.93	2 90
1920	8 20	7.79	7.22	6 64	5.58	3 28
1921	8 35	7 93	7 35	6 76	5.68	3 34
1922	7 08	6 80	6 37	5.81	4.89	3.12
1923	7.24	7 02	6 70	6.26	5 55	4.20
1924	6 83	6 69	6 49	5.94	5.19	3.89
1925	6.27	6.18	6.02	5.52	5.14	4.70
1926	5 87	5 78	5 64	5 17	4 81	4 40
1927	5.48	5 40	5.26	4 82	4.49	4 11
1928	5 48	5 40	5.26	4.82	4 49	4 11
1929	5 90	5 87	5 72	5.25	4.90	4.48
1930	5 90	5 81	5 66	5 19	4.84	4.42
1931	7.62	7.51	7.32	6 71	6.25	5.72
1932	9 30	8.93	8 37	7.63	6 42	4 09
1933	7 76	7.45	6 98	6.36	5 35	3.41
1934	6.32	6.07	5 75	5.12	4.36	2 91
1935	5.75	5 52	5 23	4 66	3 97	2.64
1936	4 77	4.58	4 34	3 86	3.29	1.96
1937	5 03	4.83	4.58	4.07	3.47	2.06
1938	5 80	5.57	5.28	4.70	4 00	2.38
1939	4.96	4 76	4 51	4 02	3 42	2.03
1940	4.75	4 54	4.23	3 13	2 45	1.62
1941	4.33	3.77	3 42	2.25	1 78	1 39
1942	4.28	3 34	2 82	1.80	1.33	0 73
1943	3 91	2 94	2.47	1.53	1 10	0 47
1944	3 61	2.71	2 24	1.37	0.90	0.36
1945	3 29	2.47	2.07	1.25	0.82	0.33
1946	3.05	2 41	2.06	1.34	0.96	0 53
1947	3.24	2 56	2 19	1 42	1.06	0 56
1948	3 47	2.89	2 80	2 31	1.67	1 27
1949	3 42	2 85	2.76	2.28	1 65	1 25
1950	3 24	2 68	2 59	2.12	1 50	1 12
1951	3 41	2.71	2.65	2.08	1.36	0 92
1952	3.52	2.74	2.65	2.04	1.20	0 88
1953	3.73	2 90	2.81	2 16	1.27	0.93
1954	3.51	2.81	2.74	2.18	1.44	0.98

* See Table 7.

CHART 6

MARKET YIELDS OF MOODY'S Aaa CORPORATE BONDS COMPARED WITH THEIR MARGINAL AFTER-TAX YIELDS TO INDIVIDUALS WITH SELECTED NET INCOMES, 1919–1954

(married, with two dependent children)



Source Table 23

sum of all farm, residential, and commercial mortagge debt and state and local government securities then outstanding.[23] It would seem that, in the whole complex of forces operating upon the disposition of individuals to invest in fixed-interest securities in this period, the deterrent influence of sharply reduced take-home yields was less important than the sum of the other forces

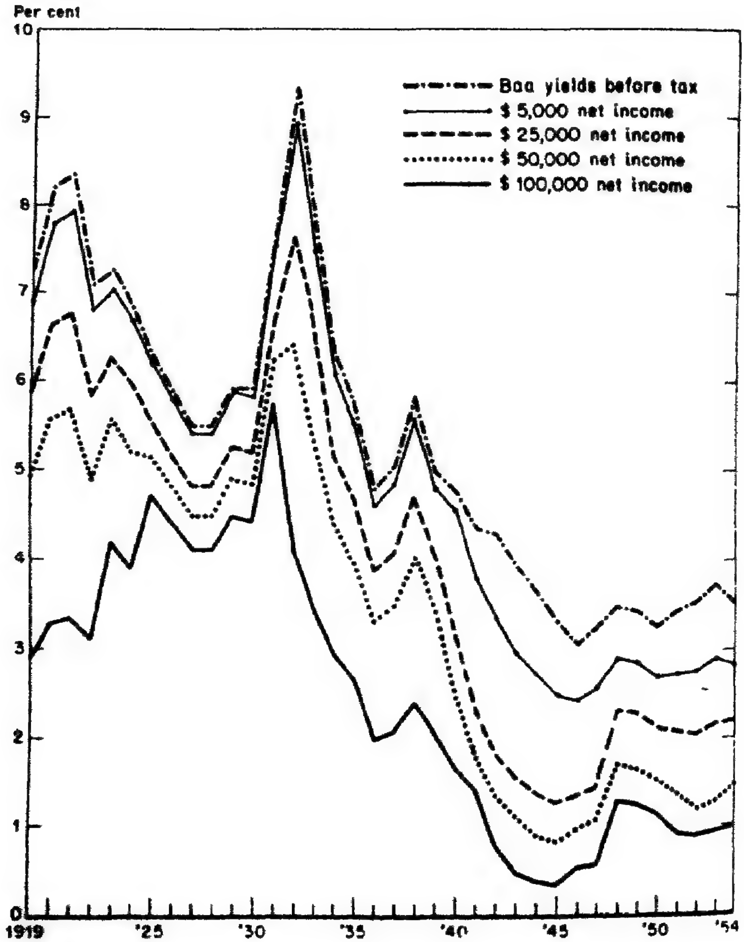
The "other" forces included the exceptional ones arising out of World War II. Between mid-1939 and mid-1946, the total amount of adjusted demand deposits and currency in circulation (outside of banks) more

[23] *Survey of Current Business*, October 1950, pp 10–15

CHART 7

MARKET YIELDS OF MOODY'S Baa CORPORATE BONDS COMPARED WITH THEIR MARGINAL AFTER-TAX YIELDS TO INDIVIDUALS WITH SELECTED NET INCOMES, 1919–1954

(married, with two dependent children)



Source: Table 24

than tripled.[24] Extensive government restrictions were imposed upon business and consumer spending and upon prices. The result was that investors not only had powerful patriotic motives and swollen financial resources to invest heavily in United States government securities during the war; they had only limited opportunities in the aggregate to spend or invest otherwise.

We may note that in the case of United States savings bonds, which accounted for the bulk of individuals' direct investments in fixed-interest securities, the deterrent influence of the combination of high income tax rates and relatively low interest rates was weakened by the option given holders to postpone tax liability for the accruing interest until redemption. Further, the seeming disposition and ability of some taxpayers to avoid reporting this and other kinds of interest receipts further tended to reduce the restrictive effect, however great or little it would otherwise have been, of the higher tax rates and lower interest rates upon fixed-interest investment.

In contrast, despite their close conceptual and competitive relationship to interest rates, the yields obtainable from common stocks behaved quite differently. Stock yields moved generally upward between 1936 and 1943, both before and after allowance for personal income taxes, though with occasional important reverses, while interest rates were falling almost uninterruptedly The average yield of Moody's list of 125 representative industrial common stocks, which had been 4 per cent in 1929 and about 3.4 per cent in 1934–1936, rose to 6.4 per cent by 1942. After declining in the next four years to 3 8 per cent, it rose sharply again to reach 6.8 per cent in 1949, whence it declined to 6.5 and 6 3 in 1950 and 1951, 5.6 in 1952, 5.5 in 1953, and 4 7 in 1954. Except in 1942–1946, common stocks offered a generally widening differential in yield over high-grade bonds until after 1950 (Table 25, Chart 8). The dividend return from Moody's representative industrial common stock average had been somewhat *below* the yield of Moody's Aaa corporate bonds in 1934 and 1935 and only slightly above in 1936. By 1949 it was well over twice that offered by the bonds, and the absolute difference between the average market yields of the stocks and bonds had risen to more than four full percentage points.

Although the high level and steeply graduated scale of the personal income tax tended to narrow this difference after allowance for taxes, the spread before income tax became sufficiently wide to give common stocks a substantial advantage in current after-tax yield. A married man with two dependent children could have obtained a marginal

[24] *Federal Reserve Bulletin*, February 1948, p. 197.

TABLE 25

COMPARISON OF MOODY'S Aaa CORPORATE BOND YIELDS WITH DIVIDEND YIELDS OF MOODY'S 125 REPRESENTATIVE INDUSTRIAL COMMON STOCKS, 1929–1954

Year	Percentage yield on		Percentage excess of stock over bond yield
	Common	Aaa bonds	
1929	4.0[a]	4.7	−0.7
1930	4.9	4.6	0.3
1931	6.4	4.6	1.8
1932	7 3	5 0	2.3
1933	3.7	4.5	−0.8
1934	3.4	4.0	−0.6
1935	3 5	3.6	−0.1
1936	3 4	3 2	0 2
1937	4 8	3.3	1.5
1938	3.9	3.2	0.7
1939	3 9	3.0	0 9
1940	5.3	2.8	2 5
1941	6.3	2 8	3.5
1942	6.4	2.8	3 6
1943	4.5	2.7	1.8
1944	4.6	2.7	1.9
1945	4 0	2.6	1 4
1946	3 8	2.5	1.3
1947	5.1	2.6	2 5
1948	5.9	2.8	3 1
1949	6 8	2.7	4 1
1950	6 5	2.6	3 9
1951	6.3	2.9	3 4
1952	5 6	3.0	2 6
1953	5 5	3.2	2 3
1954	4 7	2 9	1.8

[a] Seven-month average, June-December

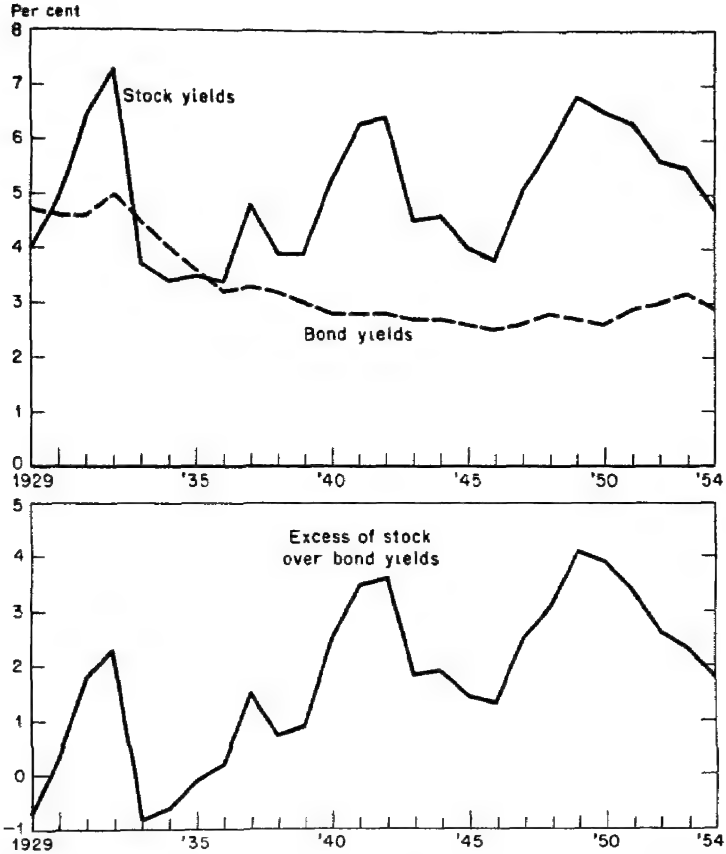
Source. Moody's Investor's Service as reprinted in *Survey of Current Business*.

after-tax yield of 5.38 per cent in 1950 from a moderate investment in Moody's list of representative industrial common stocks if his income from other sources was $5,000; 4 13 per cent, if $25.000; and 3 01 per cent, if $50,000. For the same individuals Moody's Aaa corporate bonds offered after-tax yields of only 2.16, 1.71, and 1.21 per cent respectively.

Under these circumstances, taxable fixed-interest securities doubtless

CHART 8

COMPARISON OF MOODY'S Aaa BOND YIELDS WITH DIVIDEND YIELDS OF MOODY'S 125 REPRESENTATIVE INDUSTRIAL COMMON STOCKS, 1929–1954



Source Table 25

became relatively less attractive to many individuals than competing kinds of investments, such as common stocks, rental real estate, owner-occupied houses, life insurance, tax-exempt securities, etc. For many persons who were heavily dependent upon investment income to meet their living expenses or who were for other reasons highly sensitive and responsive to the current rate of return on their investments, the

readiest of the alternatives to fixed-interest securities lay in common stocks. For some, rental real estate, such as commercial buildings and apartment houses, provided an attractive alternative because such properties—in compensation for their lesser marketability, greater risk, and greater need for personal supervision—commonly offer a higher direct return than marketable securities, as well as the important added attraction that a part of the current cash income from them is not taxable, being offset by an allowance for depreciation—an allowance which, it is usually hoped, will prove greater than the actual decline of the market value of the property. Still other persons, less concerned with current cash income or possessing special information or talent, were drawn to investments that promised rewards in forms enjoying lighter taxation, such as capital gains, oil royalties, life insurance, etc.

But the very persistence of generally high and rising stock yields in the face of much lower and declining bond yields would seem to be conclusive evidence that investors as a whole, institutional and individual, showed a strong preference for bonds as against stocks at any but substantial yield differentials. We lack reliable figures for individuals' additions to their holdings of corporate stocks during the 1940's, but the Securities and Exchange Commission has estimated that their holdings of corporate stocks, bonds, and other nongovernment securities as a whole rose by less than $2 billion during the decade.[illegible]

One factor that should logically have operated to discourage equity investment was the tendency of the increases in the level and steepness of graduation of tax rates to reduce the net yield advantage after taxes of higher-yielding risky investments over lower-yielding safer ones. When a taxpayer is subject to a 50 per cent rate on the next increment of his income (the 1950 bracket rate for a single taxpayer with net income a little over $20,000 and for a married couple filing a joint return with twice as much), the difference in marginal tax yield between a speculative stock returning 7 per cent before income tax and a safe bond yielding 3 per cent is reduced from 4 to 2 percentage points.

In cutting the "take-home" yields to 3½ and 1½ per cent, respectively (assuming an addition to income no larger than the width of one tax bracket), the tax leaves their relative relationship unchanged, but it nevertheless bears more heavily upon the risky income. A portion of the latter may be properly regarded as a kind of insurance premium or reserve required by the investor to cover the losses of capital he is

---

[illegible] Liquid savings estimates of the Securities and Exchange Commission, in *National Income Supplement, 1954, Survey of Current Business*, Table 6, p. 166.

likely to experience in the long run in connection with such investments. Instead of obtaining a continuing 7 per cent yield before taxes, he may expect in the long run to average perhaps only 5 per cent. Particularly in view of the limited deduction allowed for capital losses, it is from the *absolute* excess of yield after taxes offered by the risky over the safe security (and from possible capital gains on other investments) that the investor can hope to obtain the funds to make good his capital losses, not from the *relative* yield advantage. By reducing this absolute margin, rising or high rates of income tax, particularly if accompanied by severe limitations on the allowance for capital losses, may logically be expected to influence investors to favor safer securities or, what amounts to the same thing, to insist upon compensatory increases in the before-tax yield differentials offered by risky over safe securities.

To take an extreme illustration, consider an investor subject to a tax of 90 per cent on his next increment of income. If he choooses a risky investment with a market yield of 7 per cent over a safe one yielding 3 per cent, his after-tax yield will continue to be $2\frac{1}{3}$ times as large on the former as on the latter, but the absolute difference in after-tax yields will be only 0 4 per cent—a margin that could hardly provide significant reserves against capital losses.

Another portion of the market yield differential offered by risky over safe investments may be viewed as the compensation of the venturesome investor for the service of assuming unpopular risks and uncertainties Although the supply of this service may also be responsive to the *relationship* of the yields of risky and safe investments, apart from their absolute levels, it seems unreasonable to suppose it is not significantly responsive to the *absolute* rate of "take-home" compensation offered for it. To use our previous hypothetical example, a gross advantage in after-tax yield of *less* than 0.4 per cent (to allow something for reserves against possible capital losses) could not reasonably be expected to attract as much venturesome investment as a bigger absolute differential in yield.

Finally, there is the difficulty of recouping capital losses from ordinary income when the latter is taxed at high rates and when only severely limited offsets are allowed against it for capital losses. If the risk of loss that presumably inheres in a stock yielding 7 per cent at a time when good bonds are yielding 3 per cent should materialize, and the investor sold some or all of his shares for less than he paid for them, he would need to earn and save considerably more ordinary income than the amount of his loss to replace the latter, unless he should be fortunate enough to obtain offsetting capital gains. For

example, if a married couple filing joint returns in 1950 sustained a long-term capital loss of $28,000, and this sum was equal to their average annual income, they would need to increase their incomes by a minimum of nearly $40,000 in the aggregate during the next ten years and to save the whole increase after income taxes on it to recoup their capital loss out of ordinary income without reducing their consumption standards (assuming 1950 tax law, including a capital loss allowance against ordinary income of $1,000 each year for six years). If their average annual income approximated $50,000, and this was also the amount of their capital loss, they would require additional ordinary income aggregating about $100,000 in the next ten years to recoup their capital loss in similar fashion. Unless an investor believes he has good possibilites of offsetting possible capital losses by future capital gains, the effect of the foregoing kind of calculation is to accentuate in his mind the importance of safety of principal as against yield.

The fact that relatively high and generally rising stock yields persisted in the face of falling interest rates on high-grade obligations during the fifteen years ended in 1950 is doubtless attributable in varying degree to other influences as well as to tax considerations. Nevertheless, it seems reasonable to infer from the prevailing yields that investors made substantial allowance for the heavy personal income taxes on dividends and the limited allowance for capital losses in their appraisal of common stocks and were content during this period to accept a greatly reduced after-tax rate of return, relative to both previous levels and to current yields on equity investments, for the safety of high-grade fixed-interest securities.

The foregoing considerations apply most directly to investments that promise their rewards mainly in the form of more or less regular incomes But a considerable proportion of risky investments are expected to produce much of or all their yield in the form of capital gains, which may be loosely defined as the profits obtained by selling stocks, bonds, land, or other property not a part of the seller's stock in trade for more than they cost him Such gains are taxed at very much lower rates than ordinary income in the United States (one-half or less, provided the property is held more than six months) and most other counties, and are completely exempted from income tax in some of them.[28] In consequence, risky investments promising much of their return in this form offer investors the possibility of "take-home" yields that are

[28] See Lawrence H Seltzer, with the assistance of Selma F Goldsmith and M. Slade Kendrick, *The Nature and Tax Treatment of Capital Gains and Losses* (National Bureau of Economic Research, 1951), Chaps. 1 and 10.

substantially larger than those obtainable in the form of current income from either safe or risky investments. The combination of high tax rates on ordinary income and low ones on capital gains therefore tends to discourage some forms of risky investment and to encourage others.[27]

A study by Butters, Thompson, and Bollinger, based upon interviews with several hundred active investors representing various levels of income and wealth, presents evidence that although the investment decisions of a majority of the investors interviewed were not significantly influenced by taxes, the tax structure decreased the willingness of the interviewed investors in the aggregate to make equity-type investments; that of the investors who were influenced, a greater number were moved by tax considerations to more conservative than to more venturesome investments; that a significant proportion of the investors in the highest income groups were moved to shift to more venturesome investments in an effort to obtain the preferential tax treatment afforded capital gains; and that the ability and disposition of the uppermost income groups to supply equity capital remained strong in the face of high taxes.[28]

When yields on common stocks fail to move in the same direction as yields on high-grade bonds, a part of the explanation is doubtless to be sought in the changing prospects for capital gains. When these prospects include severe uncertainties or a deteriorating business outlook, high or rising stock yields may persist in the face of low or declining interest rates, as in most of the period between 1929 and 1949. The opposite movements may occur when these prospects are regarded more favorably by investors, as was presumably the case during the rise in stock prices and decline in stock yields in the face of firming interest rates in 1949–1954.

Among the nontax influences (or those indirectly connected with taxation) that may well have contributed to the exceptionally wide and generally increasing spread between stock and bond yields after 1936 were the continuing strong growth in the resources of banks and other institutional investors and the concentration of their investment demand upon fixed-interest obligations. The investments of commercial banks, savings banks, life insurance companies, and other financial institutions are largely restricted by law, tradition, and prudential considerations to high-grade bonds and mortgages. Net additions to the

[27] *Ibid.*, Chaps. 6 and 7, and J. Keith Butters, Laurence E. Thompson, and Lynn N. Bollinger, *Effects of Taxation: Investments by Individuals* (Graduate School of Business Administration, Harvard University, 1953).

[28] Butters, Thompson and Bollinger, *op. cit.*

supply of these, except for government obligations, were small or negative during the 1930's and the first half of the 1940's. On the other hand, the funds available for investment by financial institutions increased enormously.

Commercial banks became insistent and large-scale bidders for government and other high-grade securities. Their reserves were swelled through huge gold imports in the 1930's and through large purchases of government securities by the Federal Reserve Banks during the 1940's. Yet the total loans of all commercial banks in the United States amounted to less at the end of 1945 than at the end of 1929 or even 1924. At the end of 1929 their loans had amounted to about three times their holdings of all kinds of securities; at the end of 1945 their total loans amounted to only about one-fourth of their holdings of securities. With the loan demand restricted and new issues of high-grade corporate and state and municipal bonds small, the banks had no major alternative outlet for their greatly enlarged funds other than federal government securities during the 1930's and early 1940's. Their holdings of these rose from $5 billion in 1929 to $91 billion in 1945, when they accounted for about three-fourths of the total earning assets of commercial banks. During the several years following 1945, when the demand for business and mortgage loans rose greatly, the banks were enabled to obtain additional reserves to meet this demand by selling federal securities at more or less pegged prices to the Federal Reserve System.

Similarly, the life insurance companies and other institutional investors were insistent bidders for government and other high-grade obligations during the later 1930's and most of the 1940's because of their rapidly growing resources and the small volume of alternative investment opportunities. The 49 life insurance companies reporting to the Life Insurance Association of America increased their total admitted assets from $21.4 to $43.3 billion between 1935 and 1946 Of the $21.9 billion increase, added holdings of United States government securities accounted for $17.2 billion and public utility bonds for $3.2 billion.

For banks, insurance companies, and other financial intermediaries, the bulk of the funds that poured in each business day had to be invested fairly promptly in high-grade fixed-interest obligations at whatever yields were obtainable, with little regard to the higher yields offered by common stocks and other equity investments. The downward pressure upon the yields of high-grade fixed-interest securities therefore became especially severe. Since individuals apparently also

favored fixed-interest obligations during this period, partly for various reasons suggested above, the exceptionally wide spread between bond and stock yields appears less surprising than at first.

V. FUTURE TRENDS IN PERSONAL INTEREST INCOME

A comprehensive analysis of probable future trends in personal interest income would involve detailed consideration of a large number of complex influences. In an immediate sense the amount of direct personal interest income will be determined, of course, by the volume of public and private debt held by individuals and the level and movement of interest rates. But a host of intricate and interacting causal forces will be operating behind these immediate determinants.

The amount of *public* debt available for private investment, for example, will be determined partly by unpredictable events, such as war and business depression, and partly by deliberate decisions, to be made from year to year, with respect to the level of federal, state, and local government expenditures, the extent to which the latter are to be financed by borrowing rather than current taxation, and the provision for retiring or otherwise removing debt from the investment markets—including such quasi retirements as the important amounts acquired each year by the social security and other governmental trust funds.

Changes in the volume of *private* debt will be influenced by all the varied forces that determine the amount of private new capital formation. These include the actual and expected state of business, the demands upon the physical capacity of public utility and other important capital-using industries, the degree to which important technological changes occur that create opportunities for profitable large-scale investment, the amount of residential construction made profitable by population growth, geographical shifts in population, obsolescence, and the terms of mortgage financing, etc.

But the extent to which private new capital formation will lead to growth in private debt will in turn depend upon still other factors. These include, among others, the volume of funds becoming available for investment to financial institutions through net receipts from individuals, the amounts becoming available to business corporations each year from retained earnings and depreciation charges, and the relative popularity of bonds and mortgages as against common and preferred stocks among investors.

Of the total amount of public and private debt made available for investment, the amounts to be held directly by individuals will be

affected by the competition of banks, insurance companies, pension funds and other institutional investors for debt obligations, and by the extent to which individuals are moved to use their savings to pay off home mortgages and to add to their life insurance policies and annuity and pension contracts, etc , instead of acquiring interest-yielding investments directly.

The level and movements of interest rates will be profoundly influenced, of course, by the policies of the Federal Reserve System, as well as by the myriad other factors operating upon the supply of and demand for loanable funds.

Without attempting a detailed quantitative analysis of the foregoing and other forces, it may nevertheless be useful to note broadly a few general tendencies.

1. *Long-term uptrend in total of public and private debt*

With only short-lived interruptions the total of net interest-bearing public and private debt in the United States has been increasing for many years (Table 6). It rose from $106 billion to $474 billion between 1918 and 1952, or by an average of $11 billion a year. The cumulative or compound rate of increase was about 4½ per cent a year. The only individual years in which the total did not rise were the depression years 1931–1933 and 1938, and the year 1946, when the Treasury retired $20 billion of debt with surplus funds part of which it had obtained through its Victory Loan only a few months previously.

The largest single source of increase in 1918–1952 was the expansion in the federal debt that accompanied World Wars I and II. Nevertheless the growth of private debt and that of state and local governments accounted for 45 per cent of the total debt expansion in this period The aggregate amount of nonfederal debt outstanding declined in each of the depression years, 1931–1935 and 1938, as well as in the early war years of 1942–1943, but rose in all the other years of the period.

2. *Possible decline in net federal debt*

Barring a change in traditional federal budget practice, no large increase, if any, is to be expected, except during periods of war or business depression, in the net federal debt—the amount in other hands than federal trust funds and credit agencies. On the contrary, the steady absorption of the federal debt by federal trust funds and credit agencies, averaging $2.8 billion annually in 1944–1954, reduces by a like amount the volume of federal securities that would otherwise be

available to private investors. Hence, a balanced federal budget in the legislative and administrative sense, which includes provision for additions to the reserves or the trust funds, may mean in the future a reduction of perhaps $3 billion or more a year in the net interest-bearing federal debt outstanding in private hands. In this sense the operation of the social security and other government trust funds may be regarded as having the same effect as a sinking fund of like magnitude for reducing the amount of public debt in private hands. Some who fear that this rate of debt reduction would absorb too much of taxpayers' incomes in ordinary times have proposed that Congress seek to balance only the cash budget that is, to raise only enough tax revenue each year to meet the year's actual cash expenditures, including outpayments by the social security and other trust funds, but not additions to their reserves Others regard the public debt acquired by the social security and other trust funds as not extinguished, but alive, and as representing the funded portion of the government's growing obligations under these trust funds By gradually transferring the privately held federal debt to the trust funds, the federal government lightens its interest obligations to private holders as its social security obligations increase

3 *Special stimulants aided high rate of private debt expansion after World War II*

The net interest-bearing federal debt declined by $24 billion between the ends of 1945 and 1953, most of the decline being accounted for by the 1946 redemptions noted above But other borrowers, most notably home buyers, business corporations, and state and local governments have been increasing their interest-bearing debts at a record rate since the end of World War II. Private and state and local government debt more than doubled in the eight years ended in 1953 The average annual increase in short- and long-term nonfederal interest-bearing debt in these eight years was $21 billion, and in long-term alone, $14 billion. The largest expansion in long-term obligations took place in the noncorporate mortgage debt, which rose in the eight years from $32 billion to $92 billion Corporate long-term debt advanced from $38 billion to $79 billion, the nonmortgage debt of unincorporated business, farmers, and consumers, from $20 to $53 billion; state and local government debt, from $13 to $30 billion; and farm mortgages, included above in the noncorporate mortgage debt, from $5 to $8 billion.

Clearly if any such rate of growth were to continue, the total of personal interest income, direct and indirect, would tend to rise rapidly and substantially—barring a major decline in interest rates.

The great expansion of private and state and local government debt in 1946–1953 was a natural accompaniment of the enormous upsurge in investment that then occurred, for the greatest part of additions to interest-bearing debt is normally incurred to help finance new investment. Gross private domestic investment (as measured by the Department of Commerce), which was $16.2 billion in 1929, the previous peak year, averaged $42.6 billion annually in the eight years ending in 1953, and exceeded $50 billion in four of them.

The exceptional volume of new investment and debt in the years immediately following World War II doubtless owed much to special influences, some of which have lost their force. The period benefited from the pent-up capital needs accumulated during the depression and war for all manner of buildings and durable producers' and consumers' goods, and for the replenishment of inventories at all levels of production, distribution, and consumption.

Beginning about the middle of 1950, further stimulus to investment and to borrowing was provided by large-scale orders for military materials and equipment for the Korean War. A special incentive to a wide range of new investment was added by a law permitting business enterprises to write off in five years, for tax purposes, varying proportions of the cost of all projects for capital expenditures that were adjudged by the Office of Defense Mobilization to contribute to military defense.

The then-prevailing high rates of corporate income and excess profits taxes, coupled with the expectation that these rates would be reduced substantially in a few years, lent added attractiveness to projects eligible for rapid amortization. They also gave special encouragement to borrowing to finance much of the costs, for interest expense was deductible from taxable income. Thus, a corporation subject to a marginal rate of tax of 70 per cent on any additional income would find that its interest cost on money borrowed at 4 per cent was only 1.20 per cent, after allowance for taxes saved, whereas the dividends it would have to offer on additional preferred or common stock were not deductible from taxable income. Moreover, because a firm's borrowing increased its invested capital credit under the excess profits tax, the net effect of borrowing was sometimes to reduce its total tax liability by an amount greater than the gross interest cost of the borrowed funds.[29]

---

[29] See Donald C. Miller, "Corporate Taxation and Methods of Corporate Financing," *American Economic Review*, 42 (December 1952), 839–854.

4. *Other forces including higher price level favor further debt expansion*

Since most of or all the shortages accumulated during the depression and World War II must now be presumed to have been made up, and since the bulk of the more urgent and more profitable investment projects stimulated by the Korean War and the emergency amortization certificates also must be presumed to have been completed, a slackening in the growth of nonfederal investment and debt, and even occasional temporary reversals in trend, such as have occurred before, would not be surprising

Nevertheless, substantial average annual additions to the amount of nonfederal debt outstanding seem probable for some time to come In the absence of other supporting evidence, or of evidence to the contrary, it would be reasonable to expect such additions because they would be consistent with a long-established trend. As previously noted, and as may be seen in detail in Table 6, total nonfederal interest-bearing debt rose in twenty-five of the thirty-five years 1918–1953, and nearly all the declines were concentrated in the Great Depression and World War II. This long-term trend may be interpreted as reflecting the strong tendency of our growing economy to make sizable net additions to its stocks of capital equipment and other goods in all tolerably prosperous years and, because of a variety of factors in its institutional organization, to finance a considerable fraction of the value of such additions by borrowing Nothing in the recent behavior of the economy indicates any imminent weakening in these long-run tendencies On the contrary, various specific considerations suggest continuing strength

We have noted that noncorporate residential and commercial mortgage debt rose by $60 billion in 1945–1953 and constituted the most important component of long-term debt expansion in that period The availability of FHA and VA mortgages, with their potent attractions of small down payments and long amortization periods for borrowers, and of mortgage insurance or guaranty for lenders, was of outstanding assistance in this expansion These powerful aids to residential construction and to the growth of mortgage debt are scheduled to continue In legislation enacted in 1954, the minimum down payment under FHA-insured mortgages was reduced to 5 per cent of the first $9,000 of purchase price for new houses and 10 per cent for old houses, plus 25 per cent of any excess for both new and old houses, subject to a ceiling for insured mortgages of $20,000 on one- and two-family homes, $27,500 for three-family homes, and $35,000 on four-family homes.

Other forces besides the ease of finance promise to sustain the de-

mand for new dwellings (and the creation of mortgages) at a relatively high level, though not necessarily or uninterruptedly at the record levels of recent years. These include the high rate of population growth, the continuing geographical shifts of our population to the Northwest and the South, and the country-wide shift of population from the interiors to the outskirts or suburbs of larger cities and towns.

The same population factors, accentuated by rising standards of consumption, are creating enlarged demands upon state and local governments for public works—schools, streets, highways, sewage and water systems, etc. A considerable part of such expenditures is commonly financed by borrowing. The growth and movement of population is similarly enlarging the requirements for new facilities of public utility enterprises, which characteristically finance a substantial proportion of additions to their invested capital by borrowing. Electric light and power, gas, telephone, and pipeline, and other public utility corporations accounted for nearly one-half, and, together with the railroads, for 77 per cent, of the entire $36 billion par value of corporation bonds outstanding in the United States at the beginning of 1951[18] State and local government interest-bearing debt held outside of governmental trust and sinking funds amounted to $24 billion at that time and to $31 billion at the end of 1953.

For industrial corporations as well as public utility and railroad enterprises, the rate and character of technological change will also influence importantly the volume of new investment, and, therefore, of debt creation. Two sources of change that may conceivably require heavy investment in the relatively near future are the application of atomic energy to civilian uses and the broadening application of electronic machinery to industrial processes. Expressly intended to stimulate new investment and speedier replacement of old buildings and equipment were the optional provisions for the more rapid writing-off of new buildings and equipment enacted in the Revenue Code of 1954.

Consumer installment credit, stimulated by the increasing variety of popular durable consumer goods produced and by the still-growing market for consumer paper among banks and specialized consumer finance companies, which hold the bulk of it, promises to expand at least as rapidly as disposable personal income. The amount outstanding fell from $3.2 billion to $2.5 billion in 1942–1945, when the output of automobiles and other durable consumer goods was severely restricted by wartime controls, but it increased nearly ninefold in the next nine

[18] W. Braddock Hickman, *The Volume of Corporate Bond Financing since 1900* (National Bureau of Economic Research, 1953), Table A-1, pp. 250 ff.

years. The $22.2 billion outstanding at the end of 1953 equaled 8.9 per cent of total disposable personal income in that year as compared with 5.4 per cent in 1939.

Because of the rise in the price level since 1939—retail prices approximately doubled in 1939–1953—the dollar amounts of increases in the volume of outstanding debt will tend to be much greater than would otherwise have been the case. The upward revaluation of durable old assets, such as houses and other buildings, will tend to bring about an enlargement of the debts outstanding against them, particularly as they change hands, and the higher prices of new assets, assuming other things equal, including an unchanged proportion of borrowing against their costs, will lead to larger annual additions to the dollar volume of debt that in the past. Houses that were formerly built and sold for $6,000 and mortgaged for perhaps $4,500, now cost perhaps $12,000 to build and, with FHA assistance, command mortgages as large as $10,800 each. Thousands of old houses that still carry relatively modest mortgages issued during or shortly after the Great Depression will gradually pass into the hands of new owners at greatly increased valuations and probably will become subject to much larger mortgage debts

As cities and states build additional schools and highways and make other extensions of their physical plants to meet the requirements of an expanding and shifting population and of higher standards for public services, the cost of the new facilities in dollars per physical unit will tend to be a good deal more than in the prewar period, perhaps two or more times as much (assuming no significant cost saving through technological advances), and, with an unchanged proportion of debt financing, the new facilities would entail a correspondingly greater amount of new debt Maturing state and local government debt, incurred when prices were lower, will tend for this reason, other things being equal, to be less than the new debt.

In the same way, the transmission and distribution lines, telephone cables, pipelines, and buildings and machinery required by public utilities to meet growth and regional shifts in demand for their services will all tend to cost more than formerly per physical unit, and will therefore tend to cause greater additions to debt than the same rate of physical growth previously occasioned. In some instances it is likely that current depreciation charges based upon historical cost will be found inadequate, among industrial as well as public utility enterprises, to provide for the eventual replacement of various types of capital assets, with the result that greater resort will have to be had to retained earnings and new issues of securities, including debt issues.

At the same time the borrowing power of many enterprises will be enhanced because their old debts will constitute a reduced fraction of the present value of their assets, and their debt charges, a reduced fraction of their current earning power.

5. *Increasing demand for debt securities by financial institutions*

The growth of debt is promoted not only by the existence of attractive investment opportunities for borrowers but also by increases in the power and disposition of lenders to lend. Financial institutions make an important contribution in this respect by mobilizing the savings of thousands of individuals and assuming the responsibility and risks of investing them. Favorable to a continuing growth in debt, therefore, is the expanding volume of both voluntary and contractual or compulsory saving and investing by individuals through organizations whose investments characteristically consist mainly of fixed-interest obligations.

A major example is life insurance. As estimated by the Institute of Life Insurance, the total admitted assets of life insurance companies, largely bonds and mortgages, rose by approximately $40 billion, or more than 100 per cent, between 1943 and 1953, and have been growing in the last few years by more than $5 billion a year. Some, though not all, of this increase is doubtless to be ascribed to the rise in the general price level and in the level of wages and salaries, which encouraged individuals to take out larger insurance policies; but the dollar volume of premium receipts and of funds becoming available for investment is now geared to the higher price level.

The recent and continuing growth of private pension funds in business is another example. The following figures, taken from a study prepared for the Joint Congressional Committee on the Economic Report by the National Planning Association (*Pensions in the United States*, 1952), indicate the scale of the recent expansion in private pension plans:

ESTIMATED NUMBER OF PRIVATE PENSION PLANS
AND NUMBER OF PERSONS COVERED

Year	Number of plans	Number of persons covered (millions)
1930	720	2.4
1935	1,090	2.6
1940	1,965	3.7
1945	7,425	5.6
1950	12,330	8.6
1951	14,000	9.6

In the same study it was estimated that employers and employees contributed over $2 billion in 1951 to private pension plans, from which perhaps $300 million was paid out; that contributions under government programs, other than the old-age and survivors insurance trust fund and other than under programs for the armed forces, amounted to $2.1 billion, with benefit payments approximating $1 billion; and that the net increase in the old-age and survivors insurance trust fund in 1952 approximated $2 billion. Ida C. Merriam, of the Social Security Administration, estimated that employer contributions to private pension plans in 1954 were at the rate of about $2.3 billion a year, employee contributions, $0.5 billion, current annual additions to private pension fund reserves, about $2 4 billion, and the accumulated reserves, about $17 billion.[31] Although sizable amounts of the private pension fund reserves may come to be invested in equity ownership of rental properties and in common and preferred stocks, it seems probable that a large part of these reserves, as well as all the reserves of public pension plans, will be invested in interest-bearing obligations At the end of 1954, corporate and government bonds comprised 73 percent of the noncash assets of all uninsured pension funds of U S. corporations other than banks, insurance companies, and railroads, common stocks, 19 percent, and preferred stocks and other assets, roughly 4 percent each (S.E C Release no. 1335, Oct 12, 1955)

The widespread adoption of the practice of providing in mortgage contracts for monthly installment payments on the principal, relatively rare until the establishment of the FHA, is likewise producing a large amount of contractual saving by individuals, most of it adding to the inflow of funds to institutions and to the pressure upon them to seek replacement investments Not all of these installment principal payments represent *net* saving in an economic sense, of course, because depreciation and obsolescence of the buildings must be allowed for But the equal monthly amortization payments, after the first several years, are usually larger than the concurrent loss in the economic value of the property, even when the mortgage loans are for as long as thirty years. Moreover, the amount of funds flowing to lenders is determined by the gross, not the net, saving incorporated in installment payments of principal. Because these principal repayments reduce the volume of outstanding debt by the same amounts that they supply funds for new debt, they obviously do not contribute directly to an expansion of debt. On the contrary, by themselves, they constitute a major kind of debt reduction. But the daily receipt of huge amounts of such payments by

[31] In a paper entitled "Social Security Programs and Economic Stability," presented at the Universities-National Bureau of Economic Research Conference on policies to combat depressions, May 15, 1954.

TABLE 26

INVESTMENTS OF INDIVIDUALS IN SAVINGS ACCOUNTS, UNITED STATES SAVINGS BONDS, AND LIFE INSURANCE RESERVES, 1920–1953

(Millions of Dollars)

Dec. 31	Savings accounts				Savings bonds U. S. government[e]	Reserves of life insurance companies[f]	Total	Net increase during year
	S&LA's[a]	Mutual savings banks[b]	Commercial banks[c]	Postal savings[d]				
1920	1,741	4,806	10,546	166	761	5,488	23,508	—
1921	1,965	5,541	11,079	148	652	5,893	25,278	1,770
1922	2,210	5,985	12,289	135	730	6,360	27,709	2,431
1923	2,626	6,484	13,656	135	373	6,981	30,255	2,546
1924	3,153	6,912	15,044	137	411	7,706	33,363	3,108
1925	3,811	7,349	16,314	138	376	8,592	36,580	3,217
1926	4,378	7,799	17,237	143	356	9,594	39,507	2,927
1927	5,027	8,352	18,674	153	245	10,648	43,099	3,592
1928	5,762	8,731	19,295	158	95	11,782	45,823	2,724
1929	6,237	8,797	19,165	169	—	12,801	47,169	1,346
1930	6,296	9,384	18,647	250	—	13,690	48,267	1,098
1931	5,916	9,939	15,955	613	—	14,293	46,716	−1,551
1932	5,326	9,890	12,101	915	—	14,319	42,551	−4,165
1933	4,750	9,506	10,979	1,229	—	14,613	41,077	−1,474
1934	4,458	9,670	11,992	1,232	—	15,687	43,039	1,962
1935	4,254	9,829	12,899	1,229	153	17,203	45,567	2,528
1936	4,194	10,013	13,709	1,291	475	18,736	48,418	2,851
1937	4,080	10,126	14,410	1,303	964	20,181	51,064	2,646
1938	4,077	10,235	14,427	1,286	1,442	21,512	52,979	1,915
1939	4,118	10,481	14,865	1,315	1 900	23,024	55,703	2,724
1940	4,322	10,618	15,403	1,342	2,800	24,663	59,148	3,445
1941	4,682	10,490	15,523	1,392	5,400	26,592	64,079	4,931
1942	4,941	10,621	16,056	1,459	13,400	28,734	75,211	11,132
1943	5,494	11,707	19,001	1,837	24,700	31,365	94,104	18,893
1944	6,305	13,332	23,871	2,406	36,200	34,212	116,326	22,222
1945	7,365	15,332	29,929	3,013	42,900	37,509	136,048	19,722
1946	8,548	16,813	33,447	3,379	44,200	40,713	147,100	11,052
1947	9,753	17,744	34,694	3,523	46,200	43,820	155,734	8,634
1948	10,964	18,385	34,970	3,442	47,800	47,139	162,700	6,966
1949	12,471	19,269	35,145	3,302	49,300	50,231	169,718	7,018
1950	13,978	20,002	35,200	3,035	49,600	53,630	175,445	5,727
1951	16,073	20,880	36,592	2,806	49,100	57,140	182,593	7,148
1952	19,143	22,578	39,331	2,650	49,200	61,140	194,042	11,449
1953[g]	22,823	24,345	42,001	2,466	49,300	65,500	206,435	12,393

[a] Investments in savings and loan associations including savings accounts, deposits and investment securities. *Does not* include shares pledged against mortgage loans or investments by United States government. Source: Home Loan Bank Board.

[b] Time deposits. Source: Comptroller of the Currency, 1920–1927. National Association of Mutual Savings Banks, 1928–1947; and Federal Deposit Insurance Corporation (time deposits of individuals, partnerships and corporations), 1948 to date.

[c] Time deposits of individuals, partnerships and corporations. Source: Comptroller of the Currency, 1920–1947; and Federal Deposit Insurance Corporation, 1948 to date.

institutional lenders exerts a potent, continuing pressure upon the latter to seek new loans. This pressure is favorable to freer lending and easier borrowing. And it is not unlikely that the widespread use of the monthly repayment house mortgage induces many of the home buyers to save more than they would save otherwise, for an expanding equity in a mortgaged house does not adequately serve all the purposes of individuals' saving—it is not a ready source of liquid funds, for example.

The increasing amount of saving and indirect individual investment that is compulsory under the social security system and most private pension plans and that is contractual under life insurance policies doubtless takes place to some extent at the expense of individuals' direct investments But the restricted availability of pension rights as a source of liquid funds prior to an individual's retirement renders them an inadequate substitute for direct personal investments Because of their special objectives, the same is generally true, though in smaller degree, of life insurance policies. It would not be surprising, therefore, if the effect of these institutional developments, as of the installment payment home mortgage, is to promote a net increase in personal saving and investment To the extent that such increase takes the form of additional deposits in savings or checking accounts with banks and shares in building and loan associations, the volume of funds seeking fixed-interest investments through financial intermediaries would be increased further—barring a radical change in the predominant type of investment sought by financial institutions Finally, unless material changes in law are made or large gold discoveries occur, increases in the country's supply of money to meet the needs of an expanding population and output will require substantially equal additions to the amount of debt securities purchased by the banking system

It is striking to note that individuals raised their current additions to their holdings of savings accounts, United States saving bonds, and life insurance reserves from an average of \$2.6 billion a year in the prosperous 1920's, and \$0 9 billion in the depressed 1930's, to \$11.4 billion a year in the 1940's, and \$9 2 billion a year in the four years 1950–1953, inclusive (Table 26).

---

[d] Due depositors Outstanding principal and accrued interest on certificates of deposits, outstanding savings stamps and unclaimed deposits Source Post Office Department

[e] Current redemption value of savings held by individuals at year-end War Savings Securities, 1920–1928, and United States savings bonds, 1935 to date Does not include holdings of corporations, unincorporated businesses, pension funds, etc Source Treasury Department

[f] Accumulations in United States legal reserve life insurance companies include reserves plus dividends left to accumulate, minus premium notes and policy loans Source Institute of Life Insurance.

[g] Preliminary estimates

Note Does not include savings accounts in credit unions which amounted to approximately \$1,500 million at the close of 1953

Source Operating Analysis Division, Home Loan Bank Board

Not all the expanding resources of financial intermediaries will seek fixed-interest investments. A small proportion of the funds of life insurance companies, for example, is being invested in income-producing real estate and in common stocks. Private pension funds, because they are much less subject to statutory restrictions in this regard, may eventually devote a substantial fraction of their resources to equity investments. The number and resources of investment trusts designed primarily for investments in common stocks are increasing. But it appears safe to say that the greater part of the aggregate investment resources of financial intermediaries will continue to seek fixed-interest securities for many years to come.

6. *Debt creation and prosperity*

The considerations reviewed thus far do not prove that the volume of interest-bearing debt outstanding will necessarily rise, on the average, over the next ten or fifteen years, but they offer ground for believing such a rise to be likely. The rise will almost certainly take place if our economy avoids serious depression. This is true because a good level of business activity, by creating bright prospects, itself stimulates investment and aggregate borrowing. It is also true that a substantial increase in debt commonly functions as a major stimulus to prosperity. It contributes vitally to prosperity by financing important amounts of capital expenditures without which a high level of economic activity cannot readily be achieved. Prosperity usually requires a large volume of capital expenditures both because a sizable portion of our industrial plant and labor force is more or less specialized, in the short run, for producing buildings, machinery, and other capital goods (or the equivalent in military goods), and because our population desires not to spend all its income for perishable consumer goods but to use some of it for long-lasting capital and consumer goods and to acquire claims against the future—much of them in the legal form of debts. If the capital goods segment of our economy is not active, the consumption goods industries will also not be able to operate at full throttle, for they are geared to make and sell enough of their products for us all, including those in the capital goods industries, not only for their own employees and stockholders.

In spite of the well-founded traditional concern about the dangers of a large debt both to the debtors and to society at large,[23] an expanding volume of debt as our economy grows would seem to be necessary

[23] The dangers of debt obligations arise mainly from their inflexibility. The essential purpose of a debt contract from the creditor's standpoint is to provide a margin of protection for his claims to interest income and principal against the unavoidable risks and fluctuations of ownership of tangible property or of shares in business enterprises. For this protection he is willing to accept a limited and usually

for the full functioning of our economic institutions. By the purchase of various kinds of debt securities, both directly and through financial institutions, individuals are enabled to satisfy their desire to obtain definite and protected claims against the future in return for making available a portion of their current incomes to business corporations, governments, and others; and the latter, by productive use of the borrowed funds, help to maintain the nation's incomes and employment at a high level in the present, and place themselves in a position to meet the future obligations assumed as they mature. So long as the public continues to buy additions to its holdings of life insurance, pension and annuity contracts, and other fixed claims to money, there will be a demand by financial intermediaries for additional debt securities—barring a radical change in the predominant type of investment sought by financial institutions Similarly, as previously noted, unless we have material changes in law or large gold discoveries, increases in the country's supply of money to meet the needs of a growing population and output will require substantially equal additions to the amount of debt securities held by the commercial banks.

Direct ownership of land, buildings, machinery, and other tangible assets, or of shares of common stock in business enterprises, also provides claims on the future Because such claims are not in fixed dollar amounts nor fixed in time, but vary with the incomes, dividend declarations, and capital values actually realized, downward movements in their amounts from time to time are more or less expected and do not usually give rise to difficulties as serious as those resulting from defaults of fixed-debt contracts. For this reason, the financial organization of society may be said to be less vulnerable to adverse developments when claims on the future take the form of equity ownership predominantly and fixed debts are small.

As indicated above, however, the prevailing customs and practices of our population and our financial institutions with respect to saving and investment appear to favor the creation of a growing volume of debt. Well-informed lenders may be expected in self-protection to strive to limit the total amount and the maturities of each of their individual loans to sums judged to be well within the capacity of the

---

lower rate of return. But the creditor's secured position is obtained by increasing the risks of the debtor. The latter's promises are usually unconditional as to timing, leaving him little or no room for adjusting to temporary adversity Because of the large volume of short-term debt from some business enterprises to others, and from business enterprises as a whole to the commercial banks, and because of the general tendency of creditors to try to reduce credit in troubled times, defaults can easily become contagious. Severe and often fatal difficulties may then be created for many debtors (and their creditors) who would have surmounted their troubles if given more time. If the defaults are sufficiently large and widespread, they may do severe damage to financial confidence, and thereby impede for a more or less protracted period the ready flow of fresh money savings into new investment.

borrower, and to the extent that such lending policies are followed generally, the risks are lessened. But protection against the principal hazards of a large debt structure—panics and prolonged depressions that weaken even strong debtors—is more likely to be found in public policies designed to prevent or mitigate them, notably monetary policies.

7. *Interest Rates*

While the considerations reviewed thus far offer ground for believing that the volume of interest-bearing debt is likely to rise for some years to come, such a rise need not be accompanied by an increase in direct personal interest income. Whether the latter will also rise will depend partly on the movements in the level and structure of interest rates and partly on the extent to which individuals enlarge or reduce their direct holdings of government and corporate bonds, mortgages, and savings deposits. Without venturing a strong opinion on the probable direction of near- and intermediate-term changes in interest rates, we may note that the considerations cited below appear to indicate that a change in either direction is likely to be held to moderate proportions.

We have previously noted that interest rates on new issues of corporate and governmental debt securities and market yields on old issues fell almost continuously between 1932 and 1946 (Chart 3). Beginning in the first quarter of 1946, interest rates began an irregular two-year upward movement. A renewed decline in the three following years left long-term rates modestly above their 1946 lows. Then a sharp rise in 1951 and another in the first half of 1953 brought them to the highest levels in ten years. Between the beginning of 1946 and the middle of 1953 the average market yield of all marketable long-term Treasury bonds advanced from 2.21 to 3.09 per cent; that of Moody's Aaa corporate bonds rose from 2.54 to 3.40 per cent. The Treasury Department raised the interest rates offered on United States savings bonds sold after April 30, 1952, and, for succeeding periods, on the amounts of previous issues maturing after that date that are not tendered for early cash redemption.

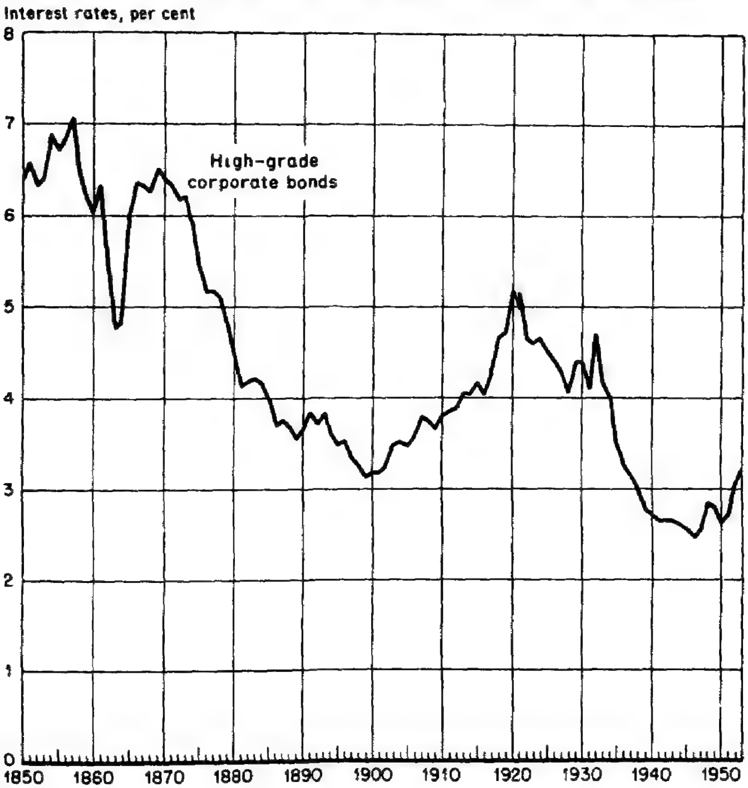
The hardening of interest rates was checked and reversed in the second half of 1953 as the combined result of a recession in business activity and of actions taken by the Federal Reserve authorities to increase the free reserves of commercial banks. In December 1954 the market yields of long-term federal and high-grade corporate bonds were roughly midway between their lows of the spring of 1946 and their highs of June 1953.

It is possible to conjecture that the net rise since 1946 may be only

the first leg of a long-term ascending movement. Chart 9, similar to one used by Raymond Goldsmith in connection with an unpublished address, depicts the course of yields on high-grade long-term corporate bonds for the last century. It will be noted that the twenty-five-year downward movement in interest rates that culminated in 1946 had been preceded by a twenty-two-year period of rising interest rates, 1899–1921, and that the latter rise, in turn, had been preceded by a thirty-year decline. Empirical parallelism suggests the possibility,

CHART 9

A CENTURY OF LONG-TERM INTEREST RATES, 1850–1953



Source *1850–1856* Leonard Ayres' average of high grade bond prices in *Business Activity and Four Price Series* (Cleveland Trust, 1932), adjusted to the level of the Macaulay series (see below) by the relation between them in 1856 *1857–1920* F R Macaulay, *The Movements of Interest Rates, Bond Yields and Stock Prices in the United States since 1856* (National Bureau of Economic Research, 1938), Table 10, col 5. *1921–1953* David Durand, *Basic Yields of Corporate Bonds, 1900–1942* (NBER, Technical Paper 3, 1942) and David Durand and Willis J Winn, *Basic Yields of Corporate Bonds, 1926–1947* (NBER, Technical Paper 6, 1947) extended to 1953 in the *Economic Almanac, 1953–54* (National Industrial Conference Board, 1953, p. 119) 40-year maturities.

though it obviously does not establish, that we began the ascending arc of long cycle of interest rate movements in 1946 and that the upward movement might continue for another decade or more.

But it would be dangerous to rely heavily upon historical parallelism in this connection. Notable changes have occurred in the central banking organization of the United States and other countries. These have diminished dependence upon gold as a source and regulator of the quantity of money and bank credit The changes have correspondingly increased the power and disposition of the central banking authorities to deliberately influence interest rates by altering the amount and cost of the reserves they make available to commercial banks.

In the use of their discretionary powers, moreover, central banking authorities have been and are still subject to strong pressures of opinion in favor of low interest rates During the Great Depression the deliberate promotion and maintenance of low interest rates became an avowed objective of many governments, first as an antidepression expedient, then as a long-run measure to combat secular unemployment. There was a striking shift of concern and emphasis on the part of many scholars and public officials from the problems of short-run cyclical instability in prices and business to the long-run difficulties of maintaining a high level of employment and output. Where depressions were once regarded as brief interludes in the steady march of economic progress, even to be welcomed in mild doses for their cathartic action upon the inefficiencies and wastes that developed during booms, they came to be regarded by many as an inveterate tendency of our economy, a ceaseless threat that needs to be combatted by all manner of direct and rearguard action Frequent central banking adjustments in both directions had commended themselves highly when credit policy was viewed as the helmsman steering the economy through a very narrow channel between inflation and depression. But when fear of a secular tendency toward depression took hold, there was concern that the effects of restrictive short-run credit policies might long outlast the occasioning circumstances and reinforce this long-run tendency. A continuous promotive policy, with low interest rates a prominent feature, therefore came to be advocated by some.

In the changed economic climate of the years 1939–1953, this type of thought receded from prominence. The 95 per cent rise in the gross national product of the United States (in constant prices) between the first and last year of this period (accompanying an increase in the civilian labor force of 15 per cent) did much to erase the former fears of long-run tendencies toward depression. At the same time, the doubling of the price level—the consumer price index rose 92 per cent and the

wholesale price index, 120 per cent—again demonstrated the reality of the danger of inflation under favoring conditions. Faith in the usefulness of a restrictive monetary policy in appropriate circumstances was restored to many who had lost it.

Nevertheless the Great Depression left an abiding legacy of public concern about maintaining a high level of employment, and a readiness to use the powers of government to this end—as evidenced by the declaration of congressional policy in the Employment Act of 1946. Monetary policy in the form of actions designed to avoid or reverse undue credit stringency is looked to by many to provide the first and most desirable method of defense against incipient business depressions. It seems reasonable to believe, therefore, that although the Federal Reserve authorities will doubtless permit market forces to move interest rates upward when loan and investment demands are heavy, and will find it necessary to deliberately restrict credit and thereby raise interest rates from time to time in order to limit speculative excesses or inflationary developments, they will have powerful motives to moderate upward movements in interest rates

Moreover, there is some evidence that the financial markets and business generally are highly sensitive to the degree of credit restraint required under present conditions to produce more than moderate absolute advances in interest rates. The greatly enlarged volume of marketable securities held by commercial banks and other institutions tends to exert an influence in this direction because even small increases in interest rates cause sizable reductions in the market value of their holdings of medium- and long-term bonds. And even moderate advances in long-term market yields tend to shrink the volume of funds available from private sources for government-sponsored lending programs, such as FHA and VA

The degree of credit restraint needed to produce a sizable advance in interest rates must be sufficient to overcome the depressive influence upon market yields of the great and growing institutional demand for high-grade obligations previously discussed—a demand that is fed to an important degree by the contractual and noncontractual current liquid savings of individuals Under these conditions, it may well be that a degree of deliberate credit restraint which causes relatively small absolute advances in interest rates may be adequate for policy purposes. It may be noted that the restrictive monetary policy pursued by the Federal Reserve authorities in the early months of 1953, in conjunction with the restrictive debt management policy of the Treasury (the offering of a long-term bond issue), had, in the words of the January 1954 *Economic Report of the President*, ". . . a more potent effect than

was generally expected." Yet the peak average monthly yield of Moody's Aaa corporate bonds (in June 1953) was only slightly more than 0.4 per cent higher than eighteen months before; and the 3.40 per cent average yield in that peak month was little more than one-half the peak levels of 1920 or 1921, and substantially below the *lowest* average annual yield of these bonds in any year between 1919 and 1935 (Table 6, Chart 3). The moderateness of the absolute rise in interest rates in 1951–1953 is the more impressive because the three-year period was one of new high records in the volume of private investment and borrowing.

Illustrating the sensitivity of long-term interest rates on the downside to a softening in the business situation and to credit-easing actions of the Federal Reserve authorities was the rapid retraction in the twelve months ended in June 1954 of about one-half of the previous seven-year net advance in long-term interest rates. The Reserve System was quick to purchase about $1 billion of Treasury bills in the open market in May and June 1953, and to announce, late in June, a reduction in the percentage reserve requirements of member banks. These two measures added somewhat more than $2 billion to the lendable reserves of the banks.

The improvement in the "quality" of large amounts of debt that results from better information about borrowers, greater mobility of investment funds, FHA insurance, and VA guarantees also operates to reduce effective interest rates Aided by telephone and telegraph communication and a nationwide network of dealers and brokers in debt securities, and lending agents for insurance companies, the institutional demand makes funds available at competitive rates to areas and borrowers, physically remote from the central money markets, that formerly paid higher rates by reason of their distance. The effective rates of interest on obligations not widely known, but otherwise of good quality, though rising and falling with other interest rates, are now less likely than formerly to make exaggerated responses to tighter money markets. Mortgages insured under the FHA or guaranteed by the Veterans' Administration are not limited to a purely local market, and even uninsured mortgages, wherever issued, may tap through local lending agents the resources of insurance companies operating on a national scale. The effect of these developments is to promise a large and continuing demand for the obligations of lesser-known borrowers and to reduce that fraction of the so-called interest rate which actually represents compensation for risk, restricted marketability, and restricted access to lenders.

On the other hand, although downward movements of interest rates are likely during business recessions, both in response to reduced de-

mands by borrowers and to deliberate action of the Federal Reserve authorities, a decline to the very low levels of the middle 1940's appears unlikely. These levels were reached in unusual conditions A combination of a depression of greater intensity and duration than had been previously experienced and huge additions to our banking reserves through imports of gold from Europe had driven interest rates down with little interruption from 1932 to 1940. Then, during most of the 1940's, direct government controls over steel, copper, and other vital materials restricted private demands for borrowed funds, while the needs of war finance encouraged an easy-money policy on the part of the Federal Reserve System and the Treasury. In the absence of a major war or a severe and prolonged depression, it appears unlikely either that Federal Reserve policy would seek a similar degree of monetary ease or that other forces would bring it about Supporting the likelihood of firmer interest rates in the intermediate future than the levels that obtained during the middle 1940's, at least, is the greatly enlarged demand for capital funds by industry, state and local governments, the housing market, and foreign countries

8. *Recent upturn in direct monetary interest income of individuals likely to continue, but to lag behind growth in total personal interest income, direct and indirect*

Even if greater changes in interest rates should occur than those here contemplated, they would affect the total of individuals' monetary interest income only gradually and far less than proportionally in the near- and intermediate-term future One reason for this is that the average rate of interest received by individual investors on their total holdings naturally moves more slowly than the market rate because the new rates are obtained only on the new components of investors' total holdings A second reason is that a considerable fraction of individuals' interest-yielding investments is in forms that are relatively slow to respond to changes in market rates of interest At the end of 1953, $91 6 billion of individuals' interest-bearing investments consisted of savings accounts with commercial banks, mutual savings banks, savings and loan associations, and the postal savings system (Table 26) An additional $49 3 billion was in United States savings bonds. Mortgages accounted for about $23 billion.[33] If we assume that individuals at the end of 1953 owned approximately the same propor-

[33] *Federal Reserve Bulletin*, June 1954, p 629 The estimate is a residual for individuals "and others" derived by subtracting from the estimated total mortgage debt outstanding the amounts held by commercial banks, trust companies, mutual savings banks, life insurance companies, building and loan associations, HOLC, FNMH, and VA Some mortgages held by other federal agencies and by business corporations are included in the holdings ascribed to "individuals and others."

tions of outstanding marketable bonds of all kinds as in Goldsmith's estimates for 1949, such securities constituted 44 per cent of their total interest-bearing investments

With the prospect of a further expansion in interest-bearing debt, and of an average interest rate not more than moderately different from the prevailing one, personal monetary interest income is likely to increase in the near- and intermediate-term future. But the growth in individuals' direct monetary receipts of interest is likely to lag behind that in total personal interest income.

Between 1943 and 1953, the direct monetary interest income of individuals and nonprofit organizations rose each year (Table 11). The annual amount moved up from $3.4 to $7.5 billion during the decade. The increase in total personal interest income, monetary and imputed, as measured by the Department of Commerce, was considerably greater absolutely, and somewhat greater proportionally, the annual amount rising from $5.8 to $13.5 billion. The latter figures, as previously noted, include not only the direct monetary interest received by individuals and nonprofit organizations, but also the value of services performed for them by financial institutions, the value being measured by the amount of the latter's net property income expended to provide such services.

Individuals will doubtless acquire directly some portion of the continuing increases in the total of mortgage obligations, public utility and other corporate bonds, and state and local obligations, although financial institutions may be expected to be the principal purchasers. A further increase in individuals' interest receipts will tend to occur as portions of their existing holdings of bonds that were acquired when interest rates were lower mature and the proceeds are reinvested—unless interest rates should lose their net rise of the last several years. Individuals would also be likely to enlarge their holdings of United States government securities, possibly at the expense of equity investments or savings deposits, if the Treasury were to refund with long-term bonds any substantial amounts of the large short-term federal debt now held mainly by banks, other financial institutions, and business corporations.

But several influences will tend to moderate the growth of individuals' direct monetary interest receipts in the near- and intermediate-term future. One of these is the volume of home mortgage repayments. For millions of relatively recent home purchasers, earlier-than-scheduled repayments of installments on the principal of their mortgage loans will offer a more attractive rate of return (in the form of reduced interest

charges) than the alternative investments open to them, and will be preferred on other grounds as well. The increase in home-mortgage indebtedness outstanding in recent years has been accompanied by a substantial and rising volume of "apparent" retirements—"apparent" because the figures include old mortgages paid off with the proceeds of new mortgage loans as well as other full payments, partial prepayments, and regular amortization. The figures for recent years are as follows:[34]

MORTGAGE LENDING ON 1- TO 4-FAMILY NONFARM HOUSES
(In Billions of Dollars)

Year	New loans made	Apparent retirements	Increase in outstandings	Outstanding at end of year
1949	11.8	7.6	4.2	37.5
1950	16.2	8 6	7 6	45.1
1951	16 4	9 6	6 8	51.9
1952	18.0	11 2	6.8	58.7
1953	19 7	12 5	7 2	65 9

Although the available data do not permit ready separation of repayments through mortgage refinancing from repayment through other means, it is clear that the latter will shortly reach sizable amounts because most of the large volume of home mortgages made in recent years calls for complete extinction of the debt through monthly payments over a period of from fifteen to thirty years.

Related to the foregoing in some respects is the possibility of heavy cash redemptions of United States savings bonds issued in the last year of World War II and the several following years, now approaching maturity. Although the Treasury offered a ten-year extension of maturity at a slightly higher interest rate to all holders of the maturing E bonds, and automatically extended the maturity of E bonds not presented for cash redemption or exchange, a large proportion of the bonds

[34] Operating Analysis Division, Home Loan Bank Board, and *Federal Reserve Bulletin*

The "New Loans Made" are actually recorded nonfarm mortgages of $20,000 or less. At present there is no reliable method of identifying the relatively small proportion of mortgages over $20,000, but these omissions appear to be approximately compensated by small nonresidential loans and junior mortgages of less than $20,000 on larger than 1- to 4-family nonfarm houses and nonresidential property Although precise estimates on home mortgage lending are not available, the recording series provides a very good gauge of the amount of permanent financing activity for all 1- to 4-family nonfarm homes Prior to 1951 when the Home Loan Bank Board prepared direct estimates of new loans on 1- to 4-family nonfarm houses, the differences in our data, for the years 1949 and 1950, exceeded the old series by 0 8 and 0.2 billion respectively, about 7 and 1 per cent for these two overlapping years

maturing in 1952 and 1953 was presented for cash redemption. One use for the funds so obtained is reduction of a home-mortgage debt.

Complicating the effect of savings bond redemptions on individuals' cash interest income is the fact that many holders will realize interest from the bonds for the first time in a legal sense when they turn them in for redemption. In other words, because most holders chose not to report their accruing discount until redemption, the act of reducing their interest-yielding assets will temporarily increase their interest income.

Conceivably, by more generous terms or more aggressive promotional efforts, the Treasury could maintain or increase the importance of savings bonds as a source of personal interest receipts It succeeded in maintaining the aggregate amount held by individuals at somewhat above $49 billion, measured by current redemption value, in 1950–1953, a period in which redemptions ran between $5 and $6 billion annually. In the same four years, it may be noted, individuals increased their savings accounts in banks, building and loan associations, and the postal savings system by $21 4 billion (Table 26).

Continuing to contest with savings bonds and other direct interest-yielding investments for the savings of individuals will be such investment intermediaries as life insurance companies and pension funds, most of whose interest earnings are not paid out as such. In the same four-year period, 1950–1953, in which individuals' holdings of savings bonds remained stationary, the reserves of life insurance companies rose by $15.3 billion, and those of private pension funds by perhaps one-half as much.

The amounts of interest income reported by individuals for income tax purposes may conceivably increase much more than their total interest receipts in the intermediate-term future. Because of the important degree of underreporting of interest on tax returns, stricter enforcement procedures may result in substantial additions to the amount of personal interest reported. As in Great Britain, such procedures might conceivably include provision for the deduction at the source and remittance to the Treasury by all important payers of interest of an amount equal to the first bracket tax rate on all interest payments. Even in the absence of such specific procedures, more thorough education of taxpayers and more comprehensive examination of tax returns may yield important results.

CORRIGENDA

Readers and authors are invited to submit corrections to papers published in any previous issue. These will be published each year, in the December issue.

Anderson, R. L., THE PROBLEM OF AUTOCORRELATION IN REGRESSION ANALYSIS, Vol. 49, No 265 (*March 1954*), 113–129.

On page 119, line 18 should read $\sum_i \sum_j b_i b_j S(\Delta X_i \Delta X_j)$ instead of $\sum_i \sum_j b_i b_j S(\Delta X_i - \Delta X_j)$.

Coale, Ansley J., THE POPULATION OF THE UNITED STATES IN 1950 CLASSIFIED BY AGE, SEX, AND COLOR—A REVISION OF CENSUS FIGURES, Vol. 50, No. 269 (*March 1955*), 16–54

On page 19, line 4, read: (with female rates reduced more than the male), instead of: (with male rates reduced more than female)

Coe, Paul F., NONWHITE POPULATION INCREASES IN METROPOLITAN AREAS, Vol. 50, No 270 (*June 1955*), 283–308.

Mr Richard L. Forstall of Rand McNally and Company has kindly pointed out that, since Laredo, Texas is located in the South, the last sentence on page 305 should be corrected to state, "Only one of these 20 SMA's was located in the South "

Errata, Vol. 49, No. 268 (*December 1954*), 907

John W Tukey was erroneously listed as a co-author with Laderman and Littauer. Line 14 should read Laderman, J , Littauer, S. B , and Weiss, Lionel.

Hildreth, Clifford, POINT ESTIMATES OF ORDINATES OF CONCAVE FUNCTIONS, Vol 49, No 267 (*September 1954*), 598–619

A hastily attempted extension of the main result of this paper can be shown to be in error Fortunately, the main result and the illustrative example are unaffected by the error. Deletion of the following material will remove references to the false extension without disrupting the organization of the article.

Lines to be deleted:

p 604, line 5 starting with "or one might . . " through line 9.

p 613, lines 9 through 23

p 615, lines 7 and 8.

In addition, the following minor errors were not detected in proofreading:

p. 602, eqn. (2.7), the period at the end of this line should be deleted.

p. 604, line 2, "x" should read "$\hat{x}$."

p. 605, line 15, "by" should read "be."

p. 612, third line from bottom, "6" should read "4."
p. 613, second line from bottom, "the" should read "and."
p. 615, eqn. (4.10), "$(X. V. W)$" should read "(X, V, W)."
p. 617, twelfth line from bottom, insert "a" between "of" and "K."
p. 618, seventh line from bottom, "ϵ" should read "3."
p. 618, fourth line from bottom, the bar above the m in $\{v^{(\bar{m})}\}$ should be deleted.

Jackson, J. Edward, and Ross, Eleanor L., EXTENDED TABLES FOR USE WITH THE "G" TEST FOR MEANS, Vol. 50, No. 270 (*June 1955*), 416–433.

On page 416, line 11, read:

$$\sqrt{\frac{n\sum X^2 - (\sum X)^2}{n(n-1)}} \text{ instead of } \sqrt{\frac{n\sum X^2 - (\sum X)}{n(n-1)}}.$$

On page 418, in the title of Table I, read:

$$G_1 = \frac{|\bar{X} - \mu|}{\bar{R}} \text{ instead of } \frac{G_1 = |\bar{X} - \mu|}{\bar{R}}.$$

Kimball, Bradford F., PRACTICAL APPLICATIONS OF THE THEORY OF EXTREME VALUES, Vol. 50, No. 270 (*June 1955*), 517–528.

On page 526, line 6, read S^* the mode, instead of S^* the median

On page 528, reference [13], E. J. Gumbel's name is misspelled.

Rider, Paul R., TRUNCATED BINOMIAL AND NEGATIVE BINOMIAL DISTRIBUTIONS, Vol. 50, No 271 (*September 1955*), 877–883.

On page 879, line 1, the exponent $-k$ should be $-m$.

Stoller, David S., UNIVARIATE TWO-POPULATION DISTRIBUTION-FREE DISCRIMINATION, Vol. 49, No. 268 (*December 1954*) 770–777.

On page 772, line 10 should read $S_{N-m}^{(2)}(z) = h/(N-m)$ instead of $S_{N-m}^{(2)}(z) = h/N-m$.

On page 775, line 7 should end in $\leq$ Correspondingly, in line 11, the words "equals about" should, in both occurrences, be changed to "is less than."

It should be remarked that, as a consequence, the bias in $\hat{Q}(\zeta^*)$ for this example is not so large as originally stated.

The reasoning leading to the inequality is that

$$\begin{aligned} &\Pr\{|\hat{Q}(\zeta^*) - Q(\zeta)| > \beta\} \\ &= \Pr\{|\max_z (k/m - h/m)| > 2\beta\} \\ &\leq \Pr\{\max_z |k/m - h/m| > 2\beta\}, \end{aligned}$$

for although $[\max_s (k/m - h/m)] \geqq 0$, $[k/m - h/m]$ is not so bounded.

Lieberman, Gerald J., and Resnikoff, George J., SAMPLING PLANS FOR INSPECTION BY VARIABLES, Vol 50, No. 270 (*June 1955*), 457–516.

Replace Table V—Master Table for Normal and Tightened Inspection for Sampling Plans Based on the Average Range—with the table below.

Corrected Table V Master Table for Sampling Plans Based on the Average Range

Sample size code letter	Sample size	k factor	Acceptable Quality Levels (normal inspection)													
			.04 p^*	.065 p^*	.10 p^*	.15 p^*	.25 p^*	.40 p^*	.65 p^*	1.00 p^*	1.50 p^*	2.50 p^*	4.00 p^*	6.50 p^*	10.00 p^*	15.00 p^*
B	3	1.910	↓	↓	↓	↓	↓	↓	↓	▽	▽	7.59	18.86	26.94	33.69	40.47
C	4	2.234	↓	↓	↓	↓	↓	↓	↓	1.53	5.50	10.92	16.45	22.86	29.45	36.90
D	5	2.474	↓	↓	↓	↓	↓	↓	1.42	3.44	5.93	9.90	14.47	20.27	26.59	33.95
E	7	2.830	↓	↓	↓	↓	.28	.89	1.99	3.46	5.32	8.47	12.35	17.54	23.50	30.66
F	10	2.405	↓	↓	↓	.23	.58	1.14	2.05	3.23	4.77	7.42	10.79	15.49	21.06	27.90
G	15	2.379	.061	.136	.253	.430	.786	1.30	2.10	3.11	4.44	6.76	9.76	14.09	19.30	25.92
H	25	2.358	.125	.214	.336	.506	.827	1.27	1.95	2.82	3.96	5.98	8.65	12.59	17.48	23.79
I	30	2.353	.147	.240	.366	.537	.856	1.29	1.96	2.81	3.92	5.88	8.50	12.36	17.19	23.42
J	35	2.349	.165	.261	.391	.564	.883	1.33	1.98	2.82	3.90	5.85	8.42	12.24	17.03	23.21
K	40	2.346	.160	.252	.375	.539	.842	1.25	1.88	2.69	3.73	5.61	8.11	11.84	16.55	22.38
L	50	2.342	.169	.261	.381	.542	.838	1.25	1.60	2.63	3.64	5.47	7.91	11.57	16.20	22.26
M	60	2.339	.158	.244	.356	.504	.781	1.16	1.74	2.47	3.44	5.17	7.54	11.10	15.64	21.63
N	85	2.335	.156	.242	.350	.493	.753	1.12	1.67	2.37	3.30	4.97	7.27	10.73	15.17	21.05
O	115	2.333	.153	.230	.333	.468	.718	1.06	1.58	2.25	3.14	4.76	6.99	10.37	14.74	20.57
P	175	2.333	.139	.210	.303	.427	.655	.972	1.46	2.08	2.93	4.47	6.60	9.89	14.15	19.88
Q	230	2.333	.142	.215	.308	.432	.661	.976	1.47	2.08	2.92	4.46	6.57	9.84	14.10	19.82

All AQL and table values are in percent defective.

BOOK REVIEWS

Evaluation of 1954 Field Trial of Poliomyelitis Vaccine: Summary Report. *Poliomyelitis Vaccine Evaluation Center.* University of Michigan, Ann Arbor, Michigan, 1955 Pp xiv, 51 No price. See review article on pages 1005–1013

Exercises in Theoretical Statistics. *M. G Kendall.* New York: Hafner Publishing Co., 1954. Pp. vii, 179. $3.75.

HERMAN CHERNOFF, *Stanford University*

GOOD examples in theoretical statistics are not easy to find. The author has compiled this list of 400 problems of varying grades of difficulty and suggested solutions to help in his teaching of statistics. (Many of the problems were obtained from examinations at several English universities.) This reviewer feels that the need for a source of problems was so great that the usefulness of these problems far outweighs any shortcomings the book may have. In fact, I am so enthusiastic about the appearance of such a book that I find it hard to say a harsh word about it and even feel that the errors in it have a useful function.

It may develop that after several years of use defects will be discovered which will arouse criticism and that this book will be followed by others as good or better. In the meantime Kendall should be applauded for having satisfied a very definite need.

One finds that in this book there are not many problems of general theoretical interest in inference. Most of the problems of a general nature deal with distribution theory and most of the problems in inference involve special examples. In a book of only 400 problems, however, it is debatable whether this can be considered a defect. In any case, the special examples deal with a rather wide variety of topics of interest among statisticians.

The essential prerequisite for a student working on these problems is a year's course in mathematical statistics, with a calculus prerequisite. It is also desirable that the student be acquainted with advanced calculus and matrix theory, and have more than one year's experience in statistics. These problems are very well suited for graduate students in American universities who often spend years studying advanced theory without ever hearing of cumulants, partial regression coefficients, or the method of moments, and tend to avoid any contact with specific applications.

The problems have been arranged in five main groups: Distribution Theory (100 problems), Sampling Theory (100 problems), Relationships (75 problems), Estimation and Inference (75 problems), and Time Series (50 problems).

In the first group, most of the problems involve moments and characteristic functions of specific distributions or general properties of moments. Several queueing and traffic problems are also treated. One of the problems yields an illustration of the fact that two characteristic functions can coincide over an interval without being identical.

In the group on sampling, the problems are essentially similar except that they involve distributions of statistics based on samples of size larger than one. The third group is quite similar also except that the emphasis mainly is on correlation and slightly on contingency.

The group on estimation and inference mainly consists of interesting problems in various fields where statistical theory has been applied. Readers who are not theoretical statisticians will probably be most interested in reading some of these problems.

The final group involves time series. It deals with moving averages, serial correlations, periodograms, and spectra. I would have preferred to see more examples involving runs and birth and death processes

In reviewing the book, I did some of the problems both because of their interest and to check for possible errors. The results must necessarily be biased because as the pressure of time rose, the density of problems done had to decline sharply Hence the first group was by far the most thoroughly checked. The proposed solution of problem 24 involves a rather interesting inequality which is not always correct The results claimed in problems 4 and 58 are incorrect. In a few problems the questions could have been posed more precisely or conditions were implied but not stated These are of minor importance, though For example, in (25) where it is stated that a function has a single maximum, presumably a single *relative* maximum is meant. In (28), $r>1$ is implied In (79) the condition i, j, k, l distinct is implied. In (113) the wording is poor and the expression for a variance is a random variable In (144) Σ was omitted and in (214) $u=ax+by$ is the correct definition

While the presence of errors in scientific papers is to be deplored, it is my impression that there are too many students who are perfectly capable of giving proofs for incorrect statements. Because of this I would have preferred to have some serious errors and vaguely stated problems purposely scattered throughout the book. The careful checking that the author and his friends and helpers did has been too effective for my taste.

It is difficult to quarrel with the choice of problems in such a book. I believe that the author was correct in not stressing combinatorial probability I would prefer to have more problems dealing with asymptotic resuts, general results in inference and some elementary problems in decision theory. On the other hand there undoubtedly were substantial reasons for not inserting such problems.

I hope that this fine work will soon be supplemented by others, for the need for good problems is great among developing theoretical statisticians

---

The Methods of Statistics, Fourth Edition *L. H C. Tippett* New York John Wiley and Sons, Inc., 1952. Pp 395 $6.00.

CHURCHILL EISENHART, *National Bureau of Standards*

THE First Edition of this book, published in 1931, filled a then urgent need for a presentation of the methods of both the Pearsonian and Fisherian schools in comparatively simple terms The First Edition was an

immediate success. The present reviewer is one of many who today can state that it was from reading and rereading Tippett's 1931 book that he first gained a real understanding and full appreciation of the New Look in Statistics, which began with the appearance of R. A. Fisher's *Statistical Methods for Research Workers.*

This Fourth Edition, like the First, is an important book. Chapters 3–12 have to do primarily with statistical inference. They constitute the bulk and the most important part of the book. Here are treated the usual elementary tests of significance and confidence-limit procedures for samples from normal, binomial, and Poisson populations. Analysis of variance, linear and nonlinear regression, simple, multiple, and partial correlation are discussed quite fully. Sampling and the principles of experimental arrangements are touched upon more briefly. Other topics considered briefly, yet instructively, include: combination of probabilities from independent tests of significance, tests for association in rank data, analysis of co-variance, orthogonal polynomials, analysis of time series, and the use of transformations to simplify statistical analyses. A fine feature of these chapters is the clear and well organized style of the writing. Some of the more complicated concepts are discussed in such a thorough and simple manner that the concepts themselves appear to be almost obvious There is a general appeal to intuition and an abundant use of pictures, which should be especially helpful to the beginner. Illustrative examples, based on actual data from biological or industrial sources, are numerous, realistic, and instructive.

From a careful study of the Introduction and Chapters 3–12, a *non-statistician* will become acquainted with ideas and techniques that are new to him, will begin to see statistics as a general method of research that can be applied to good advantage in his own field, and will come away with a working knowledge of particular methods of statistics. He will see just exactly what these particular methods aim to accomplish, what they can and cannot do, and he will begin to have some ideas on where, when, how, and to what immediate ends they might be employed to advantage in his own work or field.

The first two chapters serve to introduce frequency distributions, averages, measures of dispersion, moments, mathematical probability, and the binomial, Poisson, normal, and chi-square distributions. The exposition here is not, in the opinion of this reviewer, up to the high standard of quality maintained throughout the rest of the book. Thus, the author made no attempt to distinquish clearly between a population *parameter* and a sample function that serves as an *estimator* of the parameter until he is forced to do so, in a footnote, by the requirements of a mathematical argument on page 83. He never clearly distinquishes between an *estimator* and an *estimate*. On the other hand, the discussion of *randomness* is exceptionally well done, and the present reviewer was particularly delighted to see the explicit discussion (on pp. 60–61, 67–69) of the relation of the Poisson distribution to the exponential distribution.

In summary, this is a very readable general introduction to the methods of statistics. A scientist or engineer without formal training in statistical theory and methodology will find here nearly everything he is likely to need in the performance of his regular duties. The mathematics required is just about what he can use with ease, although he may have been exposed to more. The auxiliary tables and charts included in the Appendix are sufficient for all ordinary applications of the methods and procedures discussed. Tippett writes with the authority of one who has personally applied statistical methods with telling effect in a variety of fields, and with the skill of one from whom many others have over the years learned to do likewise. A scientist or engineer who merely peruses a chapter or two will certainly become statistically minded, and if he reads the entire book he will be considerably enriched. Professional statisticians, and others with formal training in statistical theory and methodology, will gain valuable information from a perusal of this volume on how to present statistical concepts and techniques instructively and persuasively to practical men in other fields.

Elementary Statistics. *John M Howell and Ben K Gold* Dubuque, Iowa: Wm C Brown Company, 1954. Pp v, 154 $3 00 Paper

FRED C ANDREWS, *The University of Nebraska*

ACCORDING to the authors, this book is an attempt at satisfying a need for an elementary statistics textbook that would (1) "emphasize the fundamental principles and techniques which are applicable to all fields wherein modern statistics is useful," (2) "be intelligible to the student who has had no more mathematics than ordinary first year algebra," and (3) "Its length should not be too imposing." Taking these requirements in reverse order we see that the book has 142 pages of text divided into 9 chapters and should certainly be about the amount of material for a one semester course. The mathematical level of the book is that of first year algebra. Of course no mathematical proofs or derivations are presented. There is a chapter giving the necessary arithmetic, which would be very helpful. For the most part, facility with formulas is developed as the book proceeds. The topics considered are those of rather wide application. After chapters of introduction and arithmetic, the table of contents reads "Measurements—Quantitative Data"; "Attributes—Qualitative Data"; "The Normal Distribution"; "Estimation, Confidence Limits, and Significance Tests for Large Sample Sizes"; "The Chi-square Distribution"; "Correlation"; "Descriptive Statistics".

The authors stress heavily the computational techniques of the topics considered. In each chapter there are ample exercises and a review examination complete with answers. The application of a formula is generally demonstrated by the use of real data. It is disappointing to find that the problems are not adequately connected by a flowing discussion. The brevity of text is the main fault of the book. The emphasis on computation and minimum of explanation gives the book the appearance of a statistical recipe

collection rather than an introduction to statistical reasoning. The tex resembles a set of classroom notes intended for review use rather than ma terial intended for initial learning. I believe that it would be difficult t learn more than formulas by independent reading of this book alone.

This volume can be recommended as a reasonable problem source. It spiral binding and flexible cover make it easy to use. Also its modest pric is attractive. I do believe; however, that if principles of statistical reasonin are to be presented along with computational procedures, from this book it will be necessary to devise a rather complete set of accompanying le tures.

Elements of Statistics. *H. C. Fryer.* New York: John Wiley and Sons. 1954 Pp. viii, 262. $4.75.

ANDRE G. LAURENT, *Michigan State University*

FRYER's book, which is intended for students whose mathematical back ground does not go beyond college algebra, is one of the best—if not the best—introductory text in elementary statistics the reviewer knows of.

The subject matter has been restricted to a limited number of topics. (1) Population and Frequency Distributions (Chapters II and IV); (2) Elementary Probability (Chapter III); (3) Sampling from Binomial and Normal Populations (Chapters V and VI); (4) Linear Regression and Correlation (Chapter VII). The material covered is well chosen, thoroughly discussed and vividly illustrated by a broad variety of worked examples.

The author is careful to introduce progressively the fundamental concepts and the underlying principles of statistical method. Every effort is made to make them clearly understood. Though the student is provided with the routine methods fundamental to all the applications and even a few recent techniques (the *G* test is presented) the emphasis is on the basic ideas rather than on the technical aspects of statistics.

The treatment is well balanced, and avoids the pitfalls of overemphasis on one systematic method of approach. The book succeeds in not sacrificing too much rigorousness to simplicity. A few examples will show that only minor criticisms can be made. Wald's name is not mentioned in the Section devoted to history; the use of the word "normal" on pages 16 and 17 is somewhat ambiguous; on page 25, frequency polygons are used rather than histograms to represent distributions; on page 68, the assumption of independence in Bernoulli's trials is not stressed; an allusion to truncated distribution on page 108 could be easily misunderstood.

The style is direct and lively, the examples look realistic; these features prevent the text from being dull. Though the number of statistical publications is increasing each year, there is a critical shortage of books of similar quality *on that level*. This new addition to statistical literature will be especially welcome by elementary statistics teachers; it maintains the high standard of the series.

An Introduction to Statistics for the Social Sciences. *T. G. Connolly and W. Sluckin.* **New York: Hafner Publishing Co., 1953. Pp. vii, 154. $2.75.**

HOWARD RAIFFA, *Columbia University*

THE goals of the authors are clearly stated on the jacket cover of this book: "By filling a gap in the elementary literature of statistics this modest and useful book brings help to a wide variety of students whose studies demand some statistical insight, yet who have neither the equipment nor the inclination to pursue the subject deeply." The book is written for those students who have *no* knowledge of mathematics beyond simple arithmetic; however, some very elementary algebraic manipulations are occasionally employed and formulas, in general, are stated in symbolic fashion together with supporting numerical illustrations. Undoubtedly, the strong points in favor of the book are its brevity (154 small pages), its simplicity of language, its problem oriented motivation Simple illustrative examples are chosen from the social sciences. The book is well written from a literary standpoint, very carefully exposited, but from a content and technical point of view it falls far short of its desired goal. No exercises are included in the book.

The first 64 pages of the book deal with the elements of descriptive statistics and topics covered are raw data, grouping into class intervals, frequency polygons and histograms, measures of central tendency and variability, empirical cumulative distributions. These concepts are amply illustrated and precise algorithms for computation are given The authors attempt to tell the student how to select among various population indexes (e.g , mean, mode, median) but never discuss why a choice is necessary. Throughout the book there is a generous sprinkling of statements which are incorrect when taken literally (in the sense that one can give counter-examples to the propositions asserted) but which, in most cases, constitute fairly sound advice. For example, the authors assert.

"Use the arithmetic mean when.
a) the greatest reliability is required,
b) the distribution is reasonably symmetrical,
c) when subsequent statistical calculations are to be made "

Reliability at this point of the book is meant in the sense of "fluctuation of the measurement from sample to sample of the total population " Obviously, this advice is not sound if absolute deviations from the measure of central tendency is a pertinent index of "fluctuation," or if the population is of the Cauchy type.

In Chapter V (pages 64 to 77) the normal distribution curve is arrived at from an intuitive limiting analysis of Bernoulli Trials (coin-tossing) where the probability of heads is $\frac{1}{2}$. After this demonstration the authors state:

"The reader will have noticed that the assumption underlying the game of coin-tossing is that the chances of throwing heads or tails are equal. In other games of chance different conditions obtain. For instance, in dice-throwing the chance of a six or of any other of the six members, is one in

six (1:6). *We could have approached asymmetrical distributions along these lines.* There is no need, however, for our present purpose to consider skew distributions in this manner.

"It used to be thought at one time (a) that many distributions found in nature were normal, and (b) that there were special reasons why they should be normal. *It is not now thought that the second statement is true.* As regards the first statement, all that can be said is that many natural distributions approximate the normal form more or less closely." (Italics mine.)

As is well-known, the underlined statements above are incorrect, for the Central Limit Theorem furnishes a very basic special reason why many distributions are normal and this same theorem asserts that no asymmetrical distribution could have been approached along the lines chosen by the authors (indeed, the limiting distribution is normal). In a book of this size it is reasonable that the Central Limit Theorem is not discussed but it is not reasonable that authors of a statistics text are not familiar with its implications.

The latter half of the book deals with selected elementary topics of statistical inference. The authors' viewpoints on statistical inference predate the Neyman-Pearson school for no mention is made of two types of errors, power of a test, etc. The basic ingredients of the inference problem are never clearly stated and this leads the authors to repeatedly fall into such elementary traps and propagate such misconceptions as exemplified in the quote below on confidence statements:

"... Therefore at the one per cent level of confidence our limits are given by $7.61 \pm 3.50 \times 2.04$ or 7.61 ± 7.11 (approx.). This means that if we took a very large number of sets of readings, the chances are 99 in 100 that the true standard deviation (representing the measure of scatter of the readings) could be anywhere between 0.50 lb. and 14.72 lb."

Of course, what is meant is: that if repeated sets of readings are taken, and if for each set an interval is associated in the manner so carefully explained by the authors, then, almost certainly, the proportion of intervals containing the true standard deviation will approach .99; and, consequently, for the given reading, we have "confidence" .99 in the assertion that 7.61 ± 7.11 covers the true standard deviation.

In all tests of hypotheses, statements are made at the 1% and 5% levels. If experimental results are significant at the 5% and not at the 1% level, the authors conclude essentially that the issues are not clear-cut and that further experimentation is necessary. They do not say how large the next sample should be and how the total set of observations should be treated. Furthermore, they do not seem to be aware that the way the null hypotheses are formulated one can almost guarantee a significant difference at level 1% provided the sample sizes are sufficiently large.

The following topics are discussed in the inference section of the book:

"Reliability" *of estimated mean and standard deviation in small and large samples, significance of difference between means, contingency tables and* χ^2 *tests of independence,* χ^2 *test of agreement of empirical c.d.f. with normal*

distribution, scatter diagrams and the product moment correlation coefficient, reliability of correlation coefficients, biserial and tetrachoric correlation, Kendall's and Spearman's rank correlation coefficients.

Mr. Connolly is an extra-mural lecturer in social psychology, University of London; Mr. Sluckin is in the Department of Psychology, University of Durham.

Handbook of Graphic Presentation. *Calvin F. Schmid.* New York· The Ronald Press Company, 1954. Pp. vii, 316. $6.00

KENNETH W. HAEMER, *American Telephone and Telegraph Company*

THERE have been numerous attempts to bring together the published standards of statistical graphic presentation, and to fill in the empty spots (for which no "official" standards exist) with a consensus of the best current practices. This book is the most successful effort of this kind so far. The author was confronted with the serious obstacle that two of his major sources are somewhat out of date (*Time Series Charts*—the American Standard—was written in 1938; and *Standards of Presentation* issued by the Army, was written in 1946) Despite this, he has managed to do a highly creditable job of producing a reliable and helpful handbook

Audience: In his preface, the author states that "This book has been written as a working manual for all who are concerned with the clear presentation and interpretation of statistics in grafic form." Later, he suggests that it is intended primarily for students and beginning statisticians. These slightly contradictory aims are met by including some chapters of broad general interest, some of interest mainly to beginners, and some that will appeal to specialists.

Content: This is a reasonably thorough book, although like most books on this subject, it gives the most attention to the author's favorite kinds of chart and answers most fully the questions met in his major fields of interest In this case the charts explained in most detail are semi-log charts, frequency charts, and three dimensional charts, each of which receives a chapter. Maps, another of the author's favorite presentation forms, are also enthusiastically described at considerable length in a fourth chapter.

The seven other chapters discuss the following topics. principles of chart design, drafting techniques, curve and surface charts, bar and column charts, miscellaneous graphic forms, pictorial charts, and chart reproduction.

Most of the material in this book is well organized and presented. The author has done the reader a disservice, however, by titling the chapter about curve and surface charts "Rectilinear coordinate charts" and then neglecting to explain that several of the other chapters are also about charts based on the same scheme of two scales crossing at right angles.

Bar and column charts are discussed as though they were horizontal and vertical versions of the same thing. In fact, the author says "Basically they are identical." True, they are similar in appearance, but basically bars are

one-scale comparisons; columns use the two-scale rectilinear coordinat system.

The chapter on semi-logarithmic charts explains the mechanics of thi type clearly but goes overboard on the virtues and importance of relativ rate-of-change presentation. Overenthusiastic praise such as this: "In comparison with the arithmetic line chart it possesses most of the advantages without the disadvantages" is likely to start beginners and students on the wrong foot. This would have been a better chapter if the author had explained clearly that neither of these charts has any inherent advantages o disadvantages; that each is intended to serve a different purpose, and tha the disadvantage lies in trying to use one to do the work of the other.

The chapters on frequency charts and statistical maps are generall excellent for the purposes of this book. Both are developed much more fully than in any similar book. Relationship charts, however, would benefit from a fuller treatment. The discussion of pictorial charts is extremely good. The chapter about three dimensional charts is exceptionally good, the most thorough treatment of this subject that this reviewer has seen.

The final chapter outlines the major reproduction processes and emphasizes the importance of preparing charts with a specific method of reproduction in mind. This is a useful chapter; but it would be even more helpful if the space devoted to the history of reproduction processes had been given instead to further explanation of the uses and limitations of each process

Illustrations: In any book on graphic presentation, the illustrations are the proof of the author's understanding of the subject. Professor Schmid's are excellent. Except for a few exhibits borrowed from other sources, they are first rate examples of how charts and maps should be drawn and presented. The number of illustrations is ample. in addition to many complete charts and maps, there are a large number of stripped-down "thumbnail" examples showing the major features of different chart types.

Reliability: It is probably inevitable that every book will contain inconsistencies. This book contains some, but most of them are minor. For example, the author warns against the omission of zero lines yet uses several illustrations that have none; he states that zero lines should be made heavier than other rulings but neglects to do this consistently.

Most of the technical details in this book are sound and sensible. There are very few things in it that statisticians could quibble about. But there are some; typical examples. (double amount scales) "If zero values are omitted, the scales should be so adjusted that the zero lines would coincide if the scales were extended to zero." This statement rules out many of the most useful double-scale charts: those in which the scales have been adjusted to bring out the similarity in the pattern of two series without regard for the magnitude of their changes.

Style: The writing is generally excellent. On those topics where the author lets himself go, he is admirably clear and specific. Nevertheless, he has one small but persistent weakness that will disturb some readers: the habit of

stirring up interest in a topic, then dismissing it with a vague generality. Typical examples; page 162 (Frequency charts) "Logarithmic rulings may be used for both axes, or for either the X or Y axis, depending on the problem or purpose at hand", page 79 (Bar charts) "Also, in a well designed chart of this kind the bars should not be disproportionately long and narrow or short and wide"; page 181 (Organization charts) "The various names, titles, and descriptive labels are generally enclosed in squares, circles, or other shapes, which are connected by various kinds of lines." However, despite the presence of unresolved ideas, this book is well written and is easy to read and to understand.

Summary: The chief value of this book is that it brings together the major standards of statistical graphic presentation, and supplements these with information about special forms and problems not covered in the official "standards." It puts under one cover a valuable collection of graphic principles and procedures that the reader would otherwise have to search for in several volumes. Except when it leans too heavily on out-of-date sources, its recommendations follow the best current practices. It should be a valuable introduction for the beginner in graphic presentation, and a useful reference for the experienced chart maker

---

Limit Distributions For Sums of Independent Random Variables. *B V Gnedenko and A. N. Kolmogorov.* Translated from the Russian and annotated by K. L. Chung. With an Appendix by J. L Doob Cambridge, Mass : Addison-Wesley Publishing Co , 1954. Pp. ix, 264. $7.50

WASSILY HOEFFDING, *University of North Carolina*

This is a systematic treatment of the limiting forms and the asymptotic behavior of the distributions of sums of independent real-valued random variables as the number of summands tends to infinity. The work of recent decades, in which both authors played a prominent part, has revealed that the classical results such as the law of large numbers, the central limit theorem, and Poisson's limit for the binomial distribution are merely special cases of general limit theorems which give the final answers to a substantial class of problems that have arisen in this field. The book presents the resulting beautiful theory in a clear and simple form. The only other book in English which contains a similarly extensive, though less complete, treatment of this subject is Loève's recent *Probability Theory.* The translator K. L. Chung, has added explanatory notes, made a number of corrections, and incorporated some improvements from the Hungarian edition. The reader is expected to have a knowledge of calculus to the extent needed in operating with characteristic functions which are the principal tool in the proofs.

The book consists of three parts. The introductory Part I begins with the mathematical foundations of probability, based on the theory of measure, and goes in this respect somewhat beyond what is essential for the under-

standing of the main part of the book. Some basic questions raised in this first chapter are elaborated in more detail in an appendix by J. L. Doob. A discussion of distributions and characteristic functions is followed by a chapter on infinitely divisible distributions, which are of fundamental importance for the subject of this book. A random variable is said to be infinitely divisible if for every natural number n it can be represented as the sum of n independent, identically distributed random variables. An infinitely divisible distribution is the distribution of an infinitely divisible random variable. Thus the normal, the Poisson, the Cauchy, and the Gamma distributions are infinitely divisible. So is the improper distribution which assigns probability one to a single point. Paul Lévy's representation of the characteristic function of an infinitely divisible distribution puts in evidence the wide extent of this class of distributions.

In Part II the following general problem is considered. Let $S_n = X_{n1} + X_{n2} + \cdots X_{nk_n}$ be a sum of mutually independent random variables whose number k_n tends to infinity with n. Under what conditions do there exist constants A_n such that the distribution of $S_n - A_n$ tends to a limit distribution, and what is the form of the latter? Familiar special cases are the law of large numbers (one form of which states that if $k_n = n$ and $X_{nk} = X_k/n$, where $X_1, X_2, \cdots, X_n, \cdots$ are independent and identically distributed with mean μ, then S_n tends to μ in probability and hence has an improper limit distribution), the central limit theorem, and Poisson's limit theorem for "rare events." In the general case it is assumed that the summands X_{nk} are (asymptotically) infinitesimal, which means, roughly speaking, that each summand becomes negligibly small with high probability as n increases. We then have the theorem (due to Khintchine) that any infinitely divisible distribution, and no other, can be the limit distribution of $S_n - A_n$. Further theorems show how the constants A_n and the form of the limit distribution (if it exists) can be determined from the distributions of the summands X_{nk}. By imposing additional restrictions, necessary and sufficient conditions are obtained for the limit distribution to be of a particular type. It is important to note that the conditions for the convergence to a normal distribution are of a very general kind whereas for the convergence to any particular proper non-normal distribution much more special conditions are required. Thus the prominent role of the normal law as a limit distribution is not merely a historical accident.

The last chapter of Part II is concerned with the special case where $S_n = X_1 + \cdots + X_n$ is the sum of the first n terms of an infinite sequence of independent random variables. Here the problem is to find, if possible, constants $B_n > 0$ and A_n such that $(S_n/B_n) - A_n$ has a limit distribution. It is assumed that the random variables $X_1/B_n, \cdots, X_n/B_n$ are infinitesimal. The class L of possible limit distributions under these conditions, which was determined by Paul Lévy, turns out to be a proper subclass of the class of infinitely divisible distributions. The Poisson distribution does not belong to this class.

In the final Part III the further assumption is made that S_n is the sum of

the first n terms of a sequence of independent and *identically distributed random variables*. The possible limit distributions of $(S_n/B_n)-A_n$ are the so-called stable distributions, which were investigated by Khintchine and Paul Lévy. The class of stable distributions is a small subclass of the class L. The only proper stable distributions with finite variance are normal. The Cauchy distribution also is stable. If the class of limit distributions in the proper sense is thus severely restricted under the present assumptions, any infinitely divisible distribution may be the limit distribution of a *subsequence* of $(S_n/B_n)-A_n$, as is shown in a theorem of Khintchine.

Limit theorems are of interest in applications mainly insofar as the limit distribution may serve as an approximation of the distribution of a finite sum. The greater part of Chapter 8 is concerned with the rapidity of the approach of the distribution of S_n to the normal distribution. Only the case of identically distributed summands is considered. Some of the bounds of Berry and Esseen for the difference between the distribution function of S_n and the normal distribution function are derived. Asymptotic expansions of the type considered by Edgeworth are presented in the rigorous form due to Cramér and Esseen. The rest of Chapter 8 and the final Chapter 9 are concerned with local limit theorems, that is, with the approach of the probability density or the discrete frequency function of the normalized sum to the probability density of the limit distribution.

The statistician or physicist may wonder whether limit distributions of sums of independent random variables which are not of the normal, the Poisson, or the improper types are of any importance in applications. Some indications of "unusual" limit distributions which occur in applied work can be found in the preface to this book The authors express the opinion that the theory of infinitely divisible distributions and the general limit theorems connected with them will receive in time diverse applications. If such applications so far are relatively scarce, this may perhaps be due to the fact that those working in applied fields are usually not aware of the existence of these general limit theorems.

Psychological Statistics. Second Edition. *Quinn McNemar* New York· John Wiley and Sons, Inc., 1955 Pp. viii, 408 $6 00.

LEO KATZ, *Michigan State University*

THE second edition of McNemar's book is about as good as the first (see review, this Journal, vol. 44, pp 572–574). Some minor errors have been corrected, some of the omissions in the original version have been partially filled (or,at least, mentioned) and the order of presentation has been somewhat altered. The first edition, at $4 50, was a better buy.

About 95 per cent of the revision has been done with shears and pastepot. The first three chapters are unchanged except for (a) the insertion of a paragraph, attempting to define statistics and parameters, in the first chapter, (b) the insertion of five paragraphs, on the relation between frequency polygon and histogram, in the second chapter, and (c) a minor

revision of the first paragraph of the third chapter. The old chapters 4, the normal curve and probability, 5, sampling errors and statistical inference and 12, small sample methods (which previously followed correlation and contingency) are rearranged and served up with pedagogical advantage in a unit of four chapters headed distribution curves, probability and hypothesis testing, inference: continuous variables, and small sample of t technique.

As in the first edition, the naive reader is almost certain to form a set of incorrect ideas concerning inference about distribution means. He is likely to feel that the assumption of normality of distribution is about on the same level as the use of a sharp pencil—nice, but not exactly necessary. He will probably be convinced that the question of size of sample is all-important and, consequently, will abhor medium-sized samples. Dealing with "small" samples, he will use t whether he knows the variance or not.

The remainder of the book, consisting of five chapters on correlation, one on chi-square, three on analysis of variance and one of notes on sampling is slightly improved by the addition of three sections of twenty pages on models for the analysis of variance, a chapter of six pages on comparison of variabilities and a chapter of (only) four pages on distribution-free methods. Two insertions, of less than a page each, on discriminant functions and intraclass correlation, might better have been replaced by references to more adequate discussions. Two pages, added to the chapter on chi-square, discuss exact probability computations for 2×1 and 2×2 tables but fail to mention the existence of tables giving these probabilities.

The wide acceptance of the first edition and the probable duplication of that success in this second edition raise certain questions. This reviewer feels strongly that the practitioner of statistical inference must understand much more of his art than he brings to bear on a specific problem, therefore, the "cookbook" approach cannot succeed. However, many disagree with this point of view. For these pragmatically-minded individuals, the measure of the cookbook is the quality of the cake produced. This kind of evaluation of books like McNemar's seems impossible to achieve. In the absence of such evaluation, the reviewer can only assert his judgment that McNemar's book is as good as any written in the past two decades for the purpose of putting the sharpest tools of the statistical inference trade in the untrained hands of mature, but mathematically naive, students of the social sciences.

Factorial Analysis for Non-Mathematicians. *C. J. Adcock.* Melbourne, Australia: Melbourne University Press, 1954. Pp. 88. 17s.6d.

WILLIAM G. MADOW, *Stanford University*

This brief text is intended to cover "the essentials of factor analysis in simple non-mathematical terms. There seems to be a need for a book which will enable the student to get a quick insight into what is involved without laboring through any mathematical equations."

The titles of the nine chapters provide a good outline of the contents. They are: 1. The logic of factor analysis, 2. Correlation, 3. Factors, 4 The Principle of Rotation, 5. Finding simple structure, 6. Cluster-directed analysis, 7. Combined centroid and group analysis, 8. Interpreting the results, 9. The role of factor analysis. In addition, there is an appendix and a brief list for further reading.

The book is clearly written and has carefully selected illustrations. With supplementation presenting methods of factor analysis besides the centroid method, it could serve well as an introductory text in computing factors were it not for the absence of problems. I would think that a book stressing computing so much as this would contain problems whose answers appeared at the end of the book—particularly when the procedures involve as much personal decision as does rotation.

It is unfortunate that, probably because of its brevity, the book contains so little discussion of the underlying logic of factor analysis. Factor analysis attempts to determine the underlying structure of abilities or other characteristics which have the property of being latent either because they are not measurable directly or because the researcher chooses not to measure them directly. Spearman, in his beginning work, made a certain hypothesis concerning the relationship of the latent and manifest variables and this hypothesis was later generalized in the work of those who followed These hypotheses and their implications could be discussed with no more mathematics than the book already has

In order to make any estimate of the latent variables, it is necessary in some way to connect them with the manifest variables The usual way in which this is done is to assume that the manifest variables are linear combinations of the latent variables (an error of measurement variable is usually included) At this point, it turns out that these simple hypotheses just are not definite enough to permit the estimation of the factors and consequently it becomes necessary to make further assumptions that enable one to make these estimates. Such assumptions have been made by Thurstone, Hotelling, Thomson, Burt, and others. Again, some discussion of the differences among these methods might have been given

Whatever assumptions are made one can, if he wishes, rotate, but this again involves still further assumptions These assumptions might also have been discussed in non-mathematical terms.

As is well known, there are considerable errors of measurements in any psychological test Consequently, when one is factoring, the question occurs how many factors one should take. Much work has been done of recent years, in attempting to get at least a preliminary answer to this question but this work is not referred to in the current volume. But the questions must be answered practically in one way or another in each factor analysis.

I hope that with the development of high-speed computing machines, the very computational aspects to which this book is devoted, will become unimportant. There was a time when students working on their doctor's

degrees in psychology in an area in which factor analysis was the method used, spent much more time computing the factors than on the psychological aspects of their theses. Today, one can derive a full factor analysis in a rather brief time on various electronic computers by using procedures other than the centroid procedure. Thus, the oft-repeated justification for the centroid procedure namely, its relative simplicity in computation, is ceasing to exist and other methods may well take over. In a sense, this is as it should be. It is exceedingly unfortunate to have difficulties of computation determining what is computed.

Some particular comments follow. At several points on pp. 8, 9, and 30 the author sets up straw men and vigorously demolishes them. I think that whether or not one holds these opinions, it is well to give the beginner in factor analysis an opportunity to form his own judgment on the uses of mathematics and not to indoctrinate the student under the pretext of defending him. As previously said in this review, I think the discussion of the logic of factor analysis is gravely inadequate from the point of view of the non-mathematician. A little on spurious correlation might have been in order here.

It is incorrect to say, p. 30, that the first factor accounts for as much variance as possible by one factor. Not centroid but principal components factor analysis has this property.

I wish the author had discussed what one has at the end of a factor analysis. Often, psychologists seem to differ on whether factor analysis primarily organizes information in a way that the analyst finds of help in his psychological thinking or whether the factor analysis itself finds underlying factors that have some intrinsic existence. Again, some discussion of this point would have been very interesting and helpful to all.

Thus, in this reviewer's judgment, the author did a rather nice piece of work on the computing procedure for factor analysis by the centroid method but did not discuss the logic or implications of factor analysis adequately.

Decision Processes. *R. M. Thrall, C. H. Coombs and R. L. Davis, Editors.* New York: John Wiley and Sons, 1954. Pp. viii, 332. $5.00.

PATRICK SUPPES, *Stanford University*

This book is a collection of nineteen articles originating from the University of Michigan Seminar on decision processes, held at Santa Monica, California in the summer of 1952. None of the articles is meant to be a contribution to statistics proper, but almost every one has some kind of relevance to the foundations of statistical decision theory. The foundations of decision theory seem naturally to divide into three parts: the theory of utility, the theory of subjective (or objective) probability, and the theory of optimal strategy. The material in this book contributes most to the theory of utility and least to the theory of subjective probability. A good portion of the ex-

perimental results are concerned with the study of actual behavior which deviates from statistical or game-theoretic notions of optimality.

The nineteen articles are of rather uneven merit. Some such as those of Milnor and Hausner, present definitive results, while certain ones are extremely tentative, even sketchy in character and content. My general impression of this book reinforces my general impression of the theory of decision-making: the easy plums have all been plucked and a hard climb lies ahead for those in search of new fruit. Solid progress on the climb is reported in the volume under review, but there are no leaps and bounds ahead.

A good general introduction to the variety of topics treated is provided by Davis, one of the editors. The second article is an expository discussion of mathematical models and measurement theory by C. H. Coombs, Howard Raiffa, and R. M. Thrall. Although the viewpoint represented by these authors is very clearly and ably stated here, I find it hard to agree with them that any arbitrary binary relation defines a scale of measurement. My basic objection is that the classical ordinal, interval and ratio scales are *defined* by the group of numerical transformations under which the measurement results are invariant, and this group-theoretic definition does not generalize in a natural way to an arbitrary relation or even to a partial ordering.

The remaining seventeen articles are grouped in four parts. Part I is concerned with problems of individual and social choice. Article 3, by Leo A. Goodman, on methods of amalgamation, is concerned with the common elements in the theory of social choice given individual values and the theory of optimal individual behavior in situations involving uncertainty. Article 4, by John Milnor, on games against nature is the most substantial contribution in this part. Milnor shows in an elegant manner that neither the principle of LaPlace, the minimax principle, Hurwicz's generalization of minimax nor Savage's principle of minimax regret satisfies a list of intuitively desirable properties. He then characterizes each of these four principles of decision-making in terms of an appropriate subset of his list of desirable properties. The article concludes with some suggestions concerning criteria which satisfy the most important of the properties. In Article 5 Roy Radner and Jacob Marschak consider a simple game of betting on the outcome of a toss of a coin with unknown bias after an odd number of tosses, in order to show that both Hurwicz's generalization of the minimax principle and Savage's principle of minimax regret can lead to counterintuitive solutions. In Article 6 Coombs discusses the problem of social choice from a psychologist's standpoint and presents some tentative experimental results on the application of his scaling methods to the problem. In Article 7 Stefan Vail presents some tentative ideas on alternative calculi of subjective probabilities.

Part II is concerned with learning theory. In Article 8, R. R. Bush, Frederick Mosteller, and G. L. Thompson present a concise mathematical summary and some new results on the Bush-Mosteller stochastic models for learning. These models are similar in formal structure to the states-of-nature model of statistical decision theory. Some interesting trapping theorems on

asymptotic behavior are established. It is shown in an appendix that the authors' restrictive but rather natural condition on the combining of classes of responses implies that the stochastic operators representing changes in probability of response from one learning trial to another are linear. In Article 9, W. K. Estes discusses individual behavior in uncertain situations. Estes emphasizes the significance of various experimental results which indicate that when people are asked to choose between two possible events (such as one of two lights flashing) their choice patterns asymptotically approach the relative frequencies of the events. Since the optimal strategy is always to choose the more frequent event, Estes is concerned to offer explanation of the actual behavior in terms of "environmental uncertainties," such as the fact that subjects are not told whether the events occur randomly or systematically. In my opinion statistical decision theory naturally suggests two remarks about Estes' analysis. First, the experiments involved no risk to the subject. If subjects won or lost money as the result of correct or incorrect choices would the same asymptotic behavior be exhibited? Second, there is no attempt to characterize optimal behavior with respect to the a priori information given the subjects. An answer to this admittedly rather complicated problem would also be of use in analyzing the first n responses of subjects (in contrast to their asymptotic behavior). In Article 10, Merrill M. Flood presents a preliminary formulation of some ideas combining game theory and learning theory. The general orientation is very similar to that of the Bush-Mosteller learning theory.

Part III is concerned with the theory of utility and its applications. In Article 11, Gerard Debreu elegantly establishes sufficient topological conditions for a preference ordering to be representable by a numerical function (The lexicographic ordering of points of the plane provides an example of a complete ordering which is not so representable.) In Article 12, Melvin Hausner investigates the consequences of dropping the Archimedean postulate in the von Neumann and Morgenstern theory of utility. His main result is a representation theorem in terms of vector spaces. In Article 13, Thrall extends the results of Hausner in the previous article by a furthering weakening of the axioms, and then discusses various possible applications of multidimensional utility theory in game theory. These applications hinge upon having a lexicographic ordering of the vectors representing the utilities; one tries to maximize the first component, then the second, etc. In Article 14, Jacob Marschak analyzes the activities of "teams" on the basis of game theory and statistical decision theory. A *team* is defined to be a group for which the group utility function is essentially the same as the individual utility functions of the group members. Concepts of organization, communication and action are introduced to analyze team activities. The unwieldly apparatus needed for the study of optimal team behavior is to be expected in view of the many known complications of coalition behavior in n-person games. However, some simple examples are presented which argue well for the intuitive reasonableness of the general concepts defined.

In Article 15, Herbert G. Bohnert critically examines the logical structure of the utility concept. Although some of his remarks are well-taken, in choosing to analyze statements of the form "The utility of x for entities $1, 2, \cdots, n$ in the amounts $y_1, y_2, \cdots, y_n$, is equal to u utiles," he has, I think, gone amiss. Statements of this form do not represent direct observations of individual x's behavior; rather they are complicated inferences made by means of a necessarily elaborate theory of decision-making from observations of x's actual choices. The general theory of decision-making rather than isolated numerical utility statements would seem to be a better focal point of criticism.

Part IV explicitly deals with experimental studies of decision processes, although some of the previous articles also reported on experimental results (particularly Article 9). In Article 16, Paul Hoffman, Leon Festinger, and Douglas Lawrence present the results of a carefully designed experiment on the formation and dissolution of coalitions in differently motivated situations. The chief experimental variables were task importance and degree of comparability between players of the game (peer versus non-peer conditions). Although their conceptual framework does not use the theory of n-person games in any serious way, their experiment throws interesting light on some of the problems of developing a theory adequate to deal with actual behavior in competitive situations In brief their findings were that a player given an initial advantage was irrationally (from a game-theoretic standpoint) denied opportunities to form coalitions His opportunities were further reduced by an increase in the importance attached to the game. This "irrationality" of the other players was more manifest if the player with the advantage was rated as a peer rather than a non-peer. In Article 17, Coombs joins with David Beardslee to present a general model for experiments on decision-making under uncertainty Since only some very tentative experimental results are reported and most of the pages are given over to a description of the model, this article could perhaps as well have been included in Parts I or III, as Davis points out in his introductory article However, their model is experimentally oriented It is not described in a mathematically precise manner, and it is not clear that any new mathematical consequences could be derived if it were formulated sharply In Article 18, Flood makes a second contribution dealing with a pilot experiment which was partly stimulated by the results of Estes reported in Article 9. Flood's aim was to test the hypothesis that people's strategies in a "prediction" experiment are "mixed" or "pure" depending on whether they believe the events being predicted are patterned or random. The experimental results obtained did not conclusively support the hypothesis, but they were significant enough to warrant further investigation Flood tries to meet the kind of criticism concerning the question of optimality which was raised above in discussing Estes' article, but he does not examine the problem of optimality for the patterned case in any detail. In Article 19, the final article, four mathematicians, G. Kalisch, J. W. Milnor, J. Nash and E. D. Nering, report on some

interesting experiments concerning various theoretical concepts in n-person game theory: strategic equivalence of games, von Neumann-Morgenstern solutions, the Shapely value, Milnor's bounds for "reasonable" outcomes, Nash's equilibrium points. Although the experimental results are somewhat tentative and too complicated to summarise here, they represent a worthy pioneering effort in a difficult field.

Mathematics and Plausible Reasoning. Volume I, Induction and Analogy in Mathematics. Volume II, Patterns of Plausible Inference. *G. Polya.* Princeton: The Princeton University Press. London: Geoffrey Cumberlege, Oxford University Press, 1954. Volume I, pp. xvi, 280. Volume II, x, 190. $9.00; Volume I, $5.50; Volume II, $4.50.

LEONARD J. SAVAGE, *University of Chicago*

THIS excellent book has a double interest for statisticians. In the first place, I never knew a statistician who thought he knew enough mathematics for his job, and this book, especially Volume I, is effectively designed to increase the mathematical power of a remarkably wide audience, an audience including everyone, no matter how advanced, who has studied college mathematics. In the second place, nondeductive reasoning is of paramount importance to the statistician, and this book is rich in examples and insights bearing on application of such reasoning to mathematics. These applications are unlike those to which textbook statistical theory gives access, but that, I think, should enhance their value to statisticians. No one is better qualified than Polya in interest, experience, or technical or expository skill to write on nondeductive reasoning in mathematics and its role in the development of mathematical power, as this book and others he has written bear witness.

The scope and tenor of the book can be well conveyed by quoting certain passages (paragraphing mine):

> This book has various aims, closely connected with each other In the first place, this book intends to serve students and teachers of mathematics in an important but usually neglected way. Yet in a sense the book is also a philosophical essay. It is also a continuation and requires a continuation. [I, v] . . .
>
> The standards of plausible reasoning are fluid, and there is no theory of such reasoning that could be compared to demonstrative logic in clarity or would command comparable consensus. [I, v] . . . Another point concerning the two kinds of reasoning deserves our attention. Everyone knows that mathematics offers an excellent opportunity to learn demonstrative reasoning, but I contend also that there is no subject in the usual curricula of the schools that affords a comparable opportunity to learn plausible reasoning.
>
> I address myself to all interested students of mathematics of all grades and I say: Certainly, let us learn proving, but also *let us learn guessing* [I, v] . . . Mathematics is regarded as a demonstrative science. [I, vi] . . Yet mathematics in the making resembles any other human knowledge in the making. You have to guess a mathematical theorem before you prove it; you have to guess the idea of the proof before you carry through the details. You have to combine observations and follow analogies; you have to try and try again. [I, vi] . . .

I do not believe that there is a foolproof method to learn guessing At any rate, if there is such a method, I do not know it, and quite certainly I do not pretend to offer it on the following pages. The efficient use of plausible reasoning is a practical skill and it is learned, as any other practical skill, by imitation and practice. I shall try to do my best for the reader who is anxious to learn plausible reasoning, but what I can offer are only examples for imitation and opportunity for practice. In what follows, I shall often discuss mathematical discoveries, great and small. I cannot tell the true story how the discovery did happen, because nobody really knows that. Yet I shall try to make up a likely story how the discovery could have happened. I shall try to emphasize the motives underlying the discovery, the plausible inferences that led to it, in short, everything that deserves imitation. [I, vi] . . . Each chapter will be followed by examples and comments. [I, vii] . . . In order to provide (or hide) such clues with the greatest benefit to the instruction of the reader, much care has been expended not only on the contents and the form of the proposed problems, but also on their *disposition.* [I, vii] . . . In short, I tried to use all my experience in research and teaching to give an appropriate opportunity to the reader for intelligent imitation and for doing things by himself [I, vii] . The examples of plausible reasoning collected in this book may be put to another use they may throw some light upon a much agitated philosophical problem. the problem of induction. The crucial question is: Are there rules for induction? Some philosophers say Yes, most scientists think No. In order to be discussed profitably, the question should be put differently. It should be treated differently, too, with less reliance on traditional verbalisms, or on new-fangled formalisms, but in closer touch with the practice of scientists. [I, vii] . . . And so the door opens to *investigating induction inductively.* [I, viii] . . .

It is more philosophical, I think, to consider the more general idea of plausible reasoning instead of the particular case of inductive reasoning It seems to me that the examples collected in this book lead up to a definite and fairly satisfactory aspect of plausible reasoning. Yet I do not wish to force my views upon the reader In fact, I do not even state them in Vol. I; I want the examples to speak for themselves The first four chapters of Vol. II, however, are devoted to a more explicit general discussion of plausible reasoning There I state formally the patterns of plausible inference suggested by the foregoing examples, try to systematize these patterns, and survey some of their relations to each other and to the idea of probability. [I, viii] .

This work on *Mathematics and Plausible Reasoning,* which I have always regarded as a unit, falls naturally into two parts [I, viii] . . For the convenience of the student they have been issued as separate volumes. Vol I is entirely independent of Vol II, and I think many students will want to go through it carefully before reading Vol II It has more of the mathematical "meat" of the work, and it supplies "data" for the inductive investigation of induction in Vol II. [I, viii] . . I have not provided an index, since an index would tend to render the terminology more rigid than it is desirable in this kind of work. [I, ix] . . .

The present work is a continuation of my earlier book *How to Solve It.* The reader interested in the subject should read both, but the order does not matter much. [I, ix] . . Yet there are indirect references to the former book on almost every page, and in almost every sentence on some pages. [I, ix] . . . The present book is also related to a collection of problems in Analysis by G. Szegö and the author (see Bibliography). [I, ix] . . .

Some knowledge of elementary algebra and geometry may be enough to read substantial parts of the text. Thorough knowledge of elementary al-

gebra and geometry and some knowledge of analytic geometry and calculus, including limits and infinite series, is sufficient for almost the whole text and the majority of the examples and comments. [I, xi] . . . The advanced reader who skips parts that appear to him too elementary may miss more than the less advanced reader who skips parts that appear to him too complex. [I, xi] . . . Some of the problems proposed for solution are very easy, but a few are pretty hard. [I, xi] . . .

Let us note: two persons presented with the same evidence and applying the same patterns of plausible inference may honestly disagree. [II, III] . . .

The trouble with the concept of the "*credibility* of a conjecture" is that we do not know any operational definition for it. [II, 117] . . . It could be that there is no reasonable decision, no reasonable way to say which evidence is stronger than the other. This possibility is so important that it deserves a name. If there is no reasonable way to decide which evidence is stronger, E_1 or E_2, let us call E_1 ***non-comparable*** with E_2 [II, 137] . . .

Let us note this striking difference that separates the problem-solver's investigation of the workability of his plan from the inductive investigation of a mathematical or physical conjecture, or from the judicial investigation of a charge: it is the difference between a changeable, or fleeting, and a determinate, relatively well defined object. [II, 154].

The book is accurately and beautifully printed The occurrence of "equilatera" in Example 3, p. 161, is the only misprint I noted. I saw no mathematical mistakes, though there is one extramathematical point, Example (3), pp. 85–86, which is misleading at best. In this example Polya and his bank agree that Polya's balance is $331.49. How good is the evidence that they are both right? Polya attempts to adduce an answer in terms of the probability, 10^{-4}, that two five digit numbers chosen uniformly at random will agree. But, of course, there is no real possibility that Polya or his bank did choose at random. If they did each make an error, the probability that they both made the same one, or equivalent ones, may well be reckoned as much larger than 10^{-4}.

Let me end on the right note, commending the book to one and all for pleasure and profit.

Scientific Explanation: A study of the function of theory, probability and law in science. *Richard Bevan Braithwaite.* Cambridge: University Press, 1953. Pp xi+376. $8.00.

Edwin B. Wilson, *Office of Naval Research, Boston*

PROFESSIONALLY Braithwaite holds the distinguished chair of Moral Philosophy in the university of Cambridge in succession to C. D. Broad Being a philosoper he must naturally be better acquainted with philosophy than with science and cannot fail to use language carrying connotations unnatural to the scientist and to make statements about science which jar a scientist. On the other hand the professional scientist who undertakes to write on the philosophy of science without an intimate familiarity with philosophic ideology and nomenclature cannot fail to shock the philosopher by exhibitions of philosophical naiveté.

A few like Herbert Dingle or Victor Lenzen may be at home in both science

and philosophy, diverse as the philosophies and sciences are, but I am not one of them. I may note, however that the philosophies have never been unified, that despite much talk the sciences have never been unified, and that actually no single science which is alive has been unified. A philosopher writing about scientific explanation may be illuminating to his fellow philosophers but it cannot be expected that he will be any more enlightening to scientists than one of them would be to philosophers if he chose to write on philosophic explanation.

I will note first that the name of Bridgman does not occur in the index and that I can find little evidence in the book that the author places much weight upon the operational concept or upon the "philosophy" of the experimentalist or observer. "Scientific Explanation" appears to be the philosophy of the philosopher and mathematician rather than of the scientist. This is usual, I must admit, but it is onesided and how hungry scientists are for a bit of the other side may be seen from the reception accorded to Bridgman's *Logic of Modern Physics.*

The author states (p. 51). "Even if it be granted that the phrase 'hydrogen atom' has a straightforward meaning which can be understood by itself, it cannot be pretended that the words 'electron' and 'proton' can be properly understood without reference to the deductive system of physics in which the propositions expressed by means of them occur." When I was in charge of the department of physics at the Massachusetts Institute of Technology 35 years ago I was teaching that matter was electrical in nature (the neutron had not been discovered), that is, I was putting down as what the author calls the highest level hypothesis about matter that it was made up of protons and electrons, and I was doing this largely on the basis of the experimental work in the great Cavendish Laboratory. As physics, the author's statement is unintelligible to me; it may be good (oldfashioned) chemistry—the highest level hypotheses of different sciences may be different and as a matter of fact the highest level hypotheses of different parts of the same science may be contradictory. Equally unintelligible is his remark (p. 275): "On Keynes's theory inductive confirmation serves only to increase the probability of a hypothesis by multiplying it by another probability . . "—so indoctrinated am I with the mathematical notions that probabilities are proper fractions and that the product of two proper fractions is less than either. Moreover, I cannot understand either the text or the footnote (p. 247) about the exactness of the probability of a representative point falling outside a confidence belt in the discontinuous (point-binomial) case.

One should note that (p. 8) the author states· "The publicity of its data is therefore not used (as many writers would wish to use it) as the hall-mark of science " He is concerned with the private date of psychology and the social sciences. During the 15 years I was active in the Social Science Research Council we considered that such private data did not become part of science until somehow communicated to others, and then they had to be accepted with great caution. The difficulty of obtaining the communication

without influence of observer on observed and vice versa led Bohr long ago to suggest that his principle of complementarity developed to cover the inevitable interactions between the behavior of atomic objects and the measuring instruments applied to them probably applied in some form to psychological science, but I believe the author does not examine this suggestion.

Professor Braithwaite's book is divided into eleven chapters, of which the first four deal with the logic of deductive systems. This is an important aspect of scientific explanation and is well done; had it been available to me a half century ago when I was teaching symbolic logic at Yale I should have been delighted to include some of it in my course as a fruitful application.

The next three chapters deal with probability and statistics, and presumably are those which justify a review in this *Journal*. Suffice it to say that the author, while admitting his inspiration from his teacher J. M. Keynes and dedicating his book to his memory, abandons Keynes's "philosophic" approach for the mathematical. This shows the extreme logico-mathematical slant of the author and may be the best claim his book has to originality of presentation for the professional philosopher. But it goes a long way toward putting him out as a statistician; for the statistician like the scientist has to be concerned primarily with the collection and arrangement of and the reasonable inferences from observed data. Some mathematics will surely help, too much will as surely hinder him in his pursuits. I do not disqualify as bad statistical workmanship the last work of Yule, *The Statistical Study of Literary Vocabulary*, because he failed to set up a background of mathematical probability theory for determining the "significance" of his results. Really new work in science and statistics is often if not usually like that.

The antithesis between statistics and probability goes back to the beginning of probability theory, being found clearly stated in a letter of Leibniz to Jac. Bernoulli, but Leibniz's name like Bridgman's does not occur in Braithwaite's index. The point Leibniz makes is that, although the calculation of probabilities is useful, it need not be too meticulous because although Nature has her habits, she does not repeat unerringly and your calculations will not constrain her to do so, and hence you should be particularly circumspect in taking account of all possible circumstances. Most philosophers and the continental statisticians Lexis, Bortkiewicz, Charlier, and S. D. Wicksell have followed Leibniz, as did Keynes; Braithwaite does not, he follows Neyman and Wald, Morgenstern and von Neumann. Thus he gives his fellow philosophers a chance to see what the most up-to-date work can do for them, so far as he is able to make his exposition intelligible to them.

The final four chapters deal with philosophic matters: The justification of induction, Laws of nature and causality, Causal and teleological explanation, Explantion of scientific laws. Many thousands of pages have been written by philosophers on these subjects and I have read a considerable number of hundreds thereof without enough enlightenment, without finding enough agreement, to justify my commenting. This is no evidence that these chapters or the many thousands of pages previously written are of no value.

Professionally philosophy is a very ancient and technical discipline. Hairs are split very fine and a highly developed special language or special usage of language may easily mislead the non-professional, who should keep out of the fracas.

I should like, however, to add a reference to an essay on scientific method which I believe has been insufficiently cited. It is to the presidential address of T. C. Chamberlin on "The Method of Multiple Working Hypotheses" printed in the *Journal of Geology*, 5, 837–48, reprinted *Ibid.*, 39, 155–65, and again in *Scientific Monthly*, 59, 357–62. As a contribution to the discussion of scientific induction, as a scientist's statement of how one may consciously try to follow Leibniz's advice to take account of all the circumstances, it seems to me to have real merit.

One of the difficulties of following the mathematical fraternity is that one may adopt some of their conventions too literally. They talk a great deal about the danger of small samples misleading one to mistake a mere fluctuation of sampling for a reality of science, they say little about the dangers of large samples which may give a highly significant result that is largely bias and not reproducible. This difference between "errors of observation" and "constant or systematic errors" has been recognized by astronomers for a long time. And they have a way of using $P = .05$ as a standard significance level, which is all right if understood, but the businessman, the investor, the weather forecaster, the executive, or the card player who waited for that degree of significance, would be so out of the game as to be without a livelihood. Theodore Roosevelt is said to have remarked that an executive who made four right decisions out of seven was good—a "confidence" of .556 instead of .95. For practical purposes we have to take chances with our inferences, and actually no matter how meticulous our probability calculations, we have to take chances on those probabilities.

However science is always provisional and usually approximate, and thus constantly being corrected. I hope that the author's neatness of proof and definiteness of statement does not lead any reader to misconceive this fundamental characteristic of science. The author throws out some warnings, may they prove sufficient!

Transactions of the Symposium on Computing, Mechanics, Statistics, and Partial Differential Equations, Volume II. Symposium on Applied Mathematics. **Sponsored** by the American Mathematical Society and Office of Ordnance Research, U. S. Army. New York and London. Interscience Publishers, 1955. Pp. 216. $5.00.

RICHARD BELLMAN, *The RAND Corporation*

THIS volume consists of eleven papers containing invited addresses delivered at the Second Symposium in Applied Mathematics held at the University of Chicago, April 29 and 30, 1954.

The papers of particular interest to statisticians are those by H. O. Hartley and J. Neyman, dealing with technical aspects of statistical theory,

two papers on numerical analysis by M. R. Hestenes and J. Todd, and a paper by P. M. Morse on Operations Research. For those interested in probability theory there are papers by J. E. Mayer on statistical mechanics and by W. Feller on differential equations of the type occurring in diffusion theory.

The remaining papers are three in number, a paper by J. J. Stoker on stability theory, a paper by F. J. Bureau on partial differential equations, and a paper by C. Truesdell on elasticity theory.

We shall discuss the papers in the above-mentioned order, allocating the greater part of our attention to those papers which we feel will be of most interest to the readers of this journal

The paper by J. Neyman, entitled "The Problem of Inductive Inference" consists of two parts. The first part is devoted to a discussion of basic concepts of statistical theory, the frequentist versus the nonfrequentist theories of inductive inference. In particular, points of agreement and disagreement between Neyman and recent studies of R. Carnap and R. B. Braithwaite are discussed.

The second part of the paper is devoted to a detailed application of these ideas to the problem of the determination of the decay rates of certain physical particles called neutral V-particles.

The paper concludes with a brief summary of recent trends in statistical theory, and contains a bibliography of fifty-three papers.

Following this is a paper by H. O Hartley, "Some Recent Developments in the Analysis of Variance." The author limits himself to a discussion of two topics, multiple decisions and comparisons, and short cut procedures based on substitute measures of dispersion. In addition to describing his own techniques, he gives brief summaries of related techniques. There is a bibliography of forty-seven papers.

We come now to two papers on numerical analysis. The first paper by M. Hestenes, "Iterative Computational Methods" centers around the problem of developing iterative techniques based upon a minimum principle. The gradient and conjugate gradient methods are presented, and a third method is given for the problem of matrix inversion. The paper concludes with a discussion of some sources of errors, and eigenvalue problems.

John Todd's paper, "Motivation for Working in Numerical Analysis," is an interesting and encyclopedic survey of some of the difficult and important problems that arise in the supposedly routine task of computing numerical solutions. Todd discusses in turn optimal procedures for evaluating polynomials, increasing the rapidity of convergence of sequences, interpolation problems, characteristic roots of matrices, quadratures, and game theory. At the conclusion of the paper, he discusses some of the new exploratory trends in numerical analysis. There is a bibliography of seventy-five papers.

An expository article on "Operations Research" is contributed by P. M. Morse. He discusses a number of mathematical tools that have proved useful

in operational analysis and the types of problems to which they have been applied.

We consider next two papers in the domain of mathematical probability theory. In his paper "Two Unsolved Problems of Statistical Mechanics," J. E. Mayer presents two of the outstanding difficulties of the theory. The first is the curse of dimensionality, the fact that sensible models of molecular phenomena seem to run into problems involving impossible numbers of variables, say 10^{23}. The second occurs in the theory of rate processes where the difficulty is more primitive, namely one of logical formulation.

The second paper in probability theory is by W. Feller, "On Differential Operations and Boundary Conditions." The general problem is that of constructing a theory of the second order differential operator with general boundary conditions and not necessarily continuous coefficients. The connections with diffusion theory and the theory of semi-groups is discussed.

The longest paper in the volume, some fifty-five pages, is a detailed account of the Hadamard theory of the finite part of a divergent integral by F. J. Bureau, entitled, "Divergent Integrals and Partial Differential Equations." A number of applications, developed by the author, are given

The last two papers are in the field of mechanics and mathematical physics. J. J. Stoker discusses a number of different formulations of the concept of stability, and C. A. Truesdell discusses a topic in elasticity theory.

An Introduction to Linear Programming *A. Charnes, W. W. Cooper, and A. Henderson* New York: John Wiley and Sons, Inc, 1953 Pp. ix, 74 \$2 50 Paper.

F. A. Ficken, *University of Tennessee*

In recent years many problems of practical interest have been formulated so as to require an extreme value of a linear function subject to linear conditions (equalities or inequalities) on the variables. In a "transportation problem," for example, factories j ($j=1, \cdots, n$) with products p_j and warehouses α ($\alpha=1, \cdots, m$) with capacities w_α are given, along with unit shipment costs $c_{\alpha j}$ to α from j, and it is required to find shipments $s_{\alpha j} \geq 0$ such that $\Sigma_j s_{\alpha j}=w_\alpha$, $\Sigma_\alpha s_{\alpha j}=p_j$, and the total shipping cost $\Sigma_{\alpha j} c_{\alpha j} s_{\alpha j}$ shall be minimum. The term "linear programming" has come to be applied to the formulation and (more specifically) the solution of such problems.

A particular method of arranging the calculations leading to a solution, the "simplex method," seems to have conspicuous advantages over other methods. The book under review is an explanation of the simplex method from first principles; no other such explanation, so far as the reviewer knows, has been offered by a commercial publisher.

The authors point out (p 38) that a single code for the simplex method on a digital computer will handle, beyond the problem mentioned above, matrix inversion and related processes, solution of systems of linear inequalities, and solution of a zero-sum two-person game (there being in fact an intimate connection between linear programming and game theory). Other connec-

tions with statistics are also indicated. A less direct but powerful motivation for statisticians lies in the fact that consumers of statistical wisdom seem likely also to have problems which they may wish to have assessed with a view to linear programming.

The first half of the book, by Cooper and Henderson, is entitled "An Economic Introduction to Linear Programming." The simplex method is explained and illustrated in detail, with calculations, in terms of an "input-output" problem (specifically, a nut-mix problem). Various questions arising in the course of the calculations are fully discussed, and there is much valuable interpretative comment. The distinction between "slack" and "artificial" variables emerges clearly, and there is a detailed discussion of the dual problem. A useful bibliography is appended.

The second half of the book, entitled "Lectures on the Mathematical Theory of Linear Programming," consists presumably (in view of the footnote on p. 41) of notes on lectures by Charnes. Starting with the necessary facts from linear algebra and geometry and from the theory of convex polyhedra, a reasonably complete theoretical account of the simplex method is presented. The troublesome question of degeneracy is treated in detail by a perturbation method. There is a brief section on duality and one on the relaxation method of solving a system of linear inequalities. The reviewer's feeling that the second half is rather less satisfactory than the first is probably a matter of expository taste. There are a few notational anomalies and several misprints which should not, however, cause serious trouble.

The preface suggests selections of material for various classes of readers. The book has no index.

The volume suffers, in the opinion of the reviewer, from a lack of coordination between the two parts. The authors of the first half find themselves in need of standard terminology, notation, and results, and must either refer forward to the second half, as they do for "vectors," or develop the material themselves, as they do in their section on matrices. Many important topics are discussed twice. It might be more economical, and perhaps more enlightening as well, to give at the outset a rudimentary survey of linear algebra, and perhaps even a few of the special results used later, in such a way as to provide a linguistic and conceptual framework adequate for both the economic introduction and the later mathematical theory.

Some readers might find it helpful to note that the operations performed in passing (by "the algorithm") from one tableau to another are precisely those used when one has a system of linear equations and has selected a particular equation (say the α-th) and, in that equation, a particular variable λ_j whose coefficient $x_{\alpha j}$ does not vanish, and one wishes to eliminate λ_j from all other equations; one divides the equation by $x_{\alpha j}$ in order to give λ_j the coefficient unity, and then, in order to eliminate λ_j from the β-th equation, one multiplies the α-th equation by $x_{\beta j}$ and subtracts from the β-th equation. The "algorithm" becomes immediately intelligible in terms of this familiar operation.

Industrial Process Control by Statistical Methods. *John D. Heide.* New York: McGraw-Hill Book Co., 1952. Pp. ix, 297. $6.00.

EDWIN C. HARRINGTON, JR., *Monsanto Chemical Co.*

THE publication of a technical book, theoretically at least, should fill some gap in, or extend the frontiers of, our expanding knowledge. *Industrial Process Control by Statistical Methods* is quite definitely designed to fill the gap created by the operating man in the factory, who is anxious to apply the methods of SQC, but is without formal training in mathematics beyond high school algebra. For such a person, this book should have considerable appeal. The writing is lucid, the logic is simple, and the organization of material reflects the thought and the experience of one who has struggled to inculcate the concepts of SQC in the operating levels of a large industrial concern for many years.

To the practicing statistician or even to the engineer with a minimum of training in statistical quality control, the book will not be as satisfying as many which are addressed to a beginning, but intelligent audience. Except for the method of presentation (which is good) there is little which has not already been discussed elsewhere. The content, unfortunately, stops short of such important aspects of SQC as the theory of runs, the principles of sampling, and basic probability theory The author has weakened the book, in his attempt to avoid mathematical hurdles, by limiting the discussion of probability to the empirical derivation of the normal and binomial distributions. The central limit theorem is recognized but neither stated nor discussed, and such facets of applied statistics as the Tchebycheff Inequality are omitted completely. A fact which speaks for itself is that the denominator for the standard deviation is "n "

Nor can this reviewer feel that the book is, as its Preface claims, of " . . . value for college use by students planning to enter industry." There is no question that, lacking any kind of statistical training, they would find value in the book, but it would not be acceptable at college levels, nor, I think, did the author really intend it to be.

Several chapters are devoted to the psychological impact of SQC on a conservative management and labor force in an established plant Words of warning and techniques of persuasion are well taken, and these sections could be of great value to the harassed and potentially valuable engineer.

The book is not alone in covering much of this ground in low gear, and with its publication it is likely that prospective authors and publishers will consider the gap into which this book falls quite satisfactorily filled, certainly until a significantly different approach justifies further publishing effort.

Le Controle Statistique des Fabrications. *R. Cavé.* Paris: Eyrolles, 1953. Pp. 432 3950 fr.

ANDRE G. LAURENT, *Michigan State University*

IT IS the feeling of the reviewer that, in publishing the present book which is hastily written and whose presentation contrasts with the usual thoroughness of the author, Cavé had in view securing a wider diffusion of a

previous work he published three years ago in several French periodicals. The compilation of additional material has enlarged the original paper (about thirty pages of the text) to the size of a book and explains to some extent the lack of unity of the treatment.

The First Part (74 pages), "Statistical Material," introduces the reader to frequency distributions, elementary notions of probability, probability distributions, estimation and testing hypothesis. The author is struggling with a desire to be done with the work he undertook and pleads guilty as he starts: "The book will give . . . in the first part . . . definitions . . . necessary alas." The looseness of expression and the frequent inconsistencies of style (which are present throughout the book) will be resented by French readers, and they will make the text especially difficult for English-speaking readers.

The following passages provide examples of these defects: "We have studied . . . complete populations, that is, [populations] which contain the total number of individuals which constitute them." "The occurrence of a given value [of a random variable] is called a random event. One sees the analogy with the character defining a population, but, while we can freely choose the value x of a character X of a distribution, the value of a r.v. is determined by chance." (p. 37). "[If] $Pr[Z_1<a<Z_2]=1-\alpha$ · · · there are α chances that the interval · · · $[Z_1, Z_2]$ · · · will contain · · · $[a]$" (p. 75). "Fisher's theorem is satisfied" (p. 263, 303) means that it is legitimate to apply Fisher's theorem. Quoting further examples would be tedious.

Though the book is intended for the practical man, a more serious attempt could have been made to illuminate the underlying principles and the fundamental concepts of statistical methods. Nothing is said about sampling; neither Bernoulli's theorem nor its implications are mentioned; the definition of probability is the circular one, and there is no bridge between the discrete and the continuous case. The problem of estimation is introduced as follows (p. 67): "Statement of the problem—We have already defined several parameters summarizing the data of a table . . . , assuming that all the members of the population are known. But, generally the set of n measurements so obtained . . . is a sample drawn from a much larger population of N values. We shall denote the real parameters by the letters . . . we used in Chapter I, the estimations we shall be able to obtain from the measurements will be denoted by the same letters, with a prime, if they are computed, from the measurements, with the formula of definition." The introduction to testing hypotheses is even more sketchy: "Suppose one wants to test the validity of a hypothesis about a population parameter a. For instance, one wants to accept the Hypothesis H_0, if $a=a_0$ (a given value) or to test a Hypothesis H_1 against a Hypothesis H_2. The operating characteristic curve . . . is that curve giving the probability of accepting the assumed hypothesis (for instance $a=a_0$) as a function of the possible values of a." The operating characteristic curves are then classified according to the range of the parameter.

Several statements are wrong or misleading. A few examples will suffice.

"Prime numbers are normally distributed" (p. 54); "The median . . . a value with rank $n/2$" (p. 67); "The median . . . a value with rank $(n-1/2$" (p. 358); "$dF(x_i)$ · · · called elementary density" (p. 31), "A [maximum likelihood] estimate [of a] is the root of $\partial L/\partial a=0$ with $\partial^2L/\partial^2a<0$." (p. 73); "[Pearson's distributions] satisfy · · · $Y'=(a-x)Y/b_0+b_1x+b_2x^2$, x varying from α to β (roots of the denominator)" (p. 32); etc. . . .

The Second Part, (76 pages), deals with process control. Chapters V and VI present the conventional Shewart control charts. Original features are: the introduction of operating characteristic curves for control charts as functions of the proportion of defective items, clearly illustrated by convenient graphs; a realistic discussion of the frequency of sampling; and the emphasis on control charts limits based on Tchebychev's inequality. Chapter VII "New Methods of Control" reproduces the material covered in the paper previously published by Cavé. (a) A first part expresses the "leading principles of the methods." The reviewer does not agree with Cavé's assertion that a process is "under control" either if the central tendency and the dispersion of the process do not vary too much *or* if they *vary in such a way that the proportion of defective items is still acceptable* In the reviewer's opinion, the soundest feature of Shewart's contribution has been to illuminate the methodological distinction to be established between the product and the machine, the lack of ability of a process to fit the specifications and the uncontrolled systematic causes which affect the production The aim of the control chart method is not to insure the producer against an unacceptable amount of defective items but rather to protect him against the unforseeable consequences of a "lack of control." I would not trust a physician who would not treat cancer as long as the pain is endurable. (b) A second part deals with Control Charts by variables In setting up modified $\overline{X}$ charts control limits based on the specifications (for known σ), the author proposes to extend the "acceptance inspection by variables" procedures to the "process control." He must be congratulated for the very useful charts and graphs he presents. (c) A third part is concerned with the "compressed limit gauging" method. A plan is devised which minimizes the amount of inspection (for known σ) for given $\alpha=\beta$ type I and II risks. The efficiency of the method is compared with that of control charts by variables. A sequential plan is set up. Original graphs and charts are presented which greatly simplify practical application.

Part III (46 pages) deals with sampling inspection of infinite lots (the hypergeometric distribution is not mentioned in the book). It summarizes the standard material on the subject and reproduces a large part of the material already covered in Chapter VII.

The intent of Part IV, "Statistical Methods in Research" (120 pages) is to provide the engineer with standard statistical tests and techniques. Despite an unattractive typographical presentation and some of the defects mentioned above, this aim is reached. This handbook of routine methods thus treated by solved exercises (all of them from the field mechanical engineering) touches upon a broad variety of topics, including variance analysis, statistical

dependence, regression (orthogonal polynomials), and is completed by an interesting discussion of the statistical aspects of interchangeability and tolerances.

The engineer will welcome the set of tables which, at the end of the book recapitulates the definitions, the notations and the standard methods. Unfortunately it contains some inaccuracies.

Numerical tables reproduced from English and American publications complete the book. There is some carelessness in the way they are presented, as far as the layout is concerned; further, the titles have been systematically modified and, very often, they do not indicate clearly the content of the tables.

The book will not bring any substantially new material to the American reader; it will reach, in France, an audience of industrialists and engineers to whom it will provide immediately applicable methods.

Proceedings Symposium on Operations Research in Business and Industry. Kansas City, Missouri: Midwest Research Institute, 1954. Pp. iv+185, $5.00. Paper.

L. H. C. TIPPETT, *Shirley Institute, Manchester*

THIS is a collection of papers given at a conference, together with a report of the somewhat scrappy discussion following each paper. The intention of the conference was apparently to tell some 250 industrial engineers and executives what operations research is about and the whole collection belongs to the class of ephemeral literature.

There are four general papers: "History and Prospects for Operations Research", by C. West Churchman; "A Discussion of the Major Tools of Operations Research," by Edward C. Varnum; "Experimentalism, Statistical Control and Cybernetics in Industrial Management," by Sebastian B. Littauer; and "Operations Engineering, Planning and Research," by J. E. Garrett. Churchman, in discussing whether there is anything new in operations research, propounds the following interesting law of scientific growth: "Whenever potential applications of the methods of scientific research grow faster than actual applications, a new development under a new name will take place to fill the gap." Most of the tools listed briefly by Varnum are statistical. Littauer's paper analyzes the fundamentals of industrial control and draws parallels with such things as servo mechanisms. Most of the substance of Garrett's paper is a lively development of the theme that industrial experiments should be done properly and that the O. R. team are the people to see that this is done.

The remaining papers deal with particular fields of application of operations research or with applications to particular problems. They are: "Operations Research Applied to the Office," by Earl Lamm; "Operations Research—A Boon to Marketing," by J. H. Davidson; "Production and Inventory Control in a Chemical Process," by Russell L. Ackoff; "Computational Experience in Solving Linear Programming Problems," by Wm.

Orchard-Hays; "Problems of Traffic and Transportation," by William Prager; "Balancing Cost and Quality in a Department to Provide Lower Over-all Manufacturing Costs," by Sherman Kingsbury; and "Scheduling of Batched Operations in Petroleum Products Pipe Lines," by Tage A. Mortensen. Some of these papers are slight sketches, some are fairly detailed works. Finally, there is a report of a "free-for-all" general discussion in which, quite clearly, a good time was had by all.

This symposium gives the reader a good general idea of the way operations research is moving in industry at this time and the practitioner will find a few examples given in sufficient detail to interest him. The industrial statistician who has not come across operations research before will wonder what all the fuss is about, since nine-tenths of the methods and problems will be familiar to him as applied statistics. Indeed, reading this symposium he will be inclined to wonder if operations research consists in calling everything a "model"—that word is used overmuch as a piece of jargon! Operations research is more than applied statistics, but it has long been recognized that the statistician who is to work successfully in any field must be more than a statistician. However, let us reflect on Churchman's law of scientific growth, cease to bother about names, and get on with the job.

"Statistics in Medical Research," *edited by Donald Mainland*, 1954, 94 pp., Section III of Volume 6 of *Methods in Medical Research*. Chicago Year Book Publishers.

J. L. Hodges, Jr., *University of California, Berkeley*

This brief work consists of fourteen sections, ten by Mainland (introduction, chance and random sampling, the planning of investigations, the modern method of clinical trial, some undesirable effects of laboratory tradition, confidence limits, standards of significance, standard deviations and standard errors, non-metrical tests of measurement data, and consultation with a statistician) and four by Mainland and Lee Herrera (analysis in relation to planning, clinical surveys, independent individual, and sample sizes). The work is organized as a single entity and will be reviewed as such.

The material is addressed to the medical research man, and very wisely seeks to provide him with "not, primarily, more knowledge of statistical tests, but a realization of what modern biological statistics implies throughout the conduct of any type of medical investigation." A few techniques are discussed, but only to illustrate concepts and principles As one result, the book makes fascinating reading for the mathematical statistician, or the statistician working in some other field, who wishes to learn something about the particular nature of statistics in medicine.

There is a great deal of good sense in the advice given, too little of which is to be found in many textbooks and courses on statistical methodology A few quotations will indicate the flavor. "To omit randomization because one cannot see clearly how bias could occur is like trusting that glassware in

chemistry is clean because it does not look dirty." "Unless the experiment has been designed and conducted in such a way as to avoid bias, *do not apply a test of significance, or ask a statistician to do so.*" "Corrections based on experiments with other subjects are always questionable; a sound experiment is self-contained." "If there is any possibility of periodic fluctuations in the readings . . . the observation times must be arranged to eliminate bias from that source." "The method of 'alternation' (allocation of treatment *A* to the 1st patient, *B* to the 2nd, *A* to the 3rd, and so on) is not advisable. It too readily allows the manipulation of the serial order of patients by clinicians who wish to steer certain patients to 1 or other of the treatments under test." "Objectivity . . . means . . . an assessment unbiased by knowledge of the treatment that the patient is receiving." "It still seems necessary to point out that the surveying of large numbers will not at all guarantee the removal of bias." "If [treatments] *V* and *W* were in use at different parts of the period under survey, there is a very considerable risk of difference between patients who, according to the records, were apparently comparable." "Great effort at high precision in 1 part of an experiment may be wasted if the precision in another part is low." " . . . it is recommended that the emphasis should be placed, not on correlation, but on regression." "The general rule should be not to reject observations without adequate knowledge of observational and biological variation." " . . . if 1 clinical treatment is applied in 1 hospital ward and another treatment in another, however many patients are in each ward there are strictly only 2 independent individuals in the experiment." "Unless . . . pairing is based on something more than mere hope that the members of a pair would respond similarly to similar treatment, it is not an advisable technique" " . . . something may be very highly significant, i.e., very unlikely to have occurred by chance, but not be very useful in practical application." "A report should always . . . state the probability P for each result. . . . The reader can then judge the evidence for himself and apply his own standard of significance." "Whether an observer grasps fully the implications of such quantities as the standard deviations or not, it is his duty to make clear to his reader what quantities he is using." "Enlargement of sample size is affected by a law of diminishing returns with regard to precision."

There are reasonably good brief discussions of the rationale of randomization, design of clinical trials, spurious correlation and other pitfalls of clinical surveys, unintended dependence of observations, meaning of confidence limits, significance tests, and standard errors, and determination of sample size. A concluding section instructs the clinician in the abased manner proper when consulting a statistician. For example: "If the statistician points out a flaw in the research method or an invalid inference, accept his statement as unlikely to be wrong, even if you do not understand it."

There are inevitably in such a work a number of more or less minor and more or less technical points which may strike some readers as unfortunate, but to emphasise them in a review would distract attention from the

purpose of the book. This is not intended as a handbook of methods, but as a presentation of a concept of statistics which will unfortunately still be novel to many in the target audience.

(The work under review appears as one of four sections of a volume. Readers of this *Journal* may also be interested in Section I, "Some methods of studying human genetics," written by C. C. Li and edited by Antonio Ciocco. The other sections are not statistical in nature.)

Clinical vs. Statistical Prediction, *Paul E. Meehl.* Minneapolis: University of Minnesota Press, 1954. Pp. x, 149. $3 00.

DAVID L. WALLACE, *University of Chicago*

IN THIS short monograph, the author presents an unbiased survey of some theoretical arguments and empirical data in comparison of clinical and statistical methods of prediction. Statistical prediction is taken to mean (pages 15, 16) "that the prediction is arrived at by some straightforward application of an equation or table to the data . . The defining property is that no judging or inferring or weighing is done by a skilled clinician. Once the data have been gathered from whatever source and of whatever type, the prediction itself could be turned over to a clerical worker " The problem is of practical importance for (page 7) " . . therapists are in shortage . . . The clerk or statistician cannot do therapy; hence it is of the greatest importance to ascertain whether the clinician can do a better job of prediction than they can. If he cannot, we are wasting his precious time "

The book consists of somewhat disconnected discussions of various aspects of the conflict between the two methods. "The style and sequence . . . reflect my own ambivalence and real puzzlement . . . " (page vi) There is no complete resolution of the controversy or even of the conflicting arguments contained within the book, yet many problems are clarified, and many difficulties resolved.

The importance of validation of clinical (or any other) predictions is emphasized throughout (particularly in Chapters 5 and 10). Attention is given to claims that clinical predictions for single individuals are essentially different from statistical predictions. Various meanings for validation of such predictions are considered, such as the validation of the class of predictions by one clinician about patients of one sort.

A major portion of the book is spent in describing some twenty empirical studies which provide comparisons between the success of clinical and statistical predictions. Defects in the design of the studies (for this purpose) are presented along with the empirical results. Common defects are different data used in making the two predictions, the use of clearly sub-optimal procedures in the statistical predictions, and the use of the same sample for establishing a prediction formula as for validation. The over-all results showed the statistical predictions equal or superior to the clinical predictions in all but one study.

Several interesting points are raised in the general discussion of the empirical results. In relation to the practical problem of predicting, should the statistical prediction be tested against the prediction of the "best" clinician or the "average" clinician? Is the "best" clinician stably such, and can he be identified?

The author notes that the empirical studies are rather similar, and quite restricted in the nature of prediction, there being predictions of success in training or schooling, recidivism, and recovery from major psychosis. The range of alternative predictions is then usually a dichotomy or a simple scale, and the behavioral form of outcome is also rather special.

The discussion at the end of Chapter 10 on the differing nature of predictions and the difference between prediction in therapy and straight prediction links up with the lengthy discussion in Chapter 6 of the logic of clinical activity. A main contention here is that when the behavior being predicted is complex, sufficient experience to permit construction of actuarial tables will not exist, and that the clinician can sometimes arrive at conclusions through construction of some structural-dynamic hypothesis for the individual. A success frequency can be obtained for a clinician operating in this manner, but it is difficult, if not impossible, to construct a comparable statistical prediction. It appears possible that the clinical prediction will be best almost solely in those situations in which there are no statistical competitors.

In Chapter 9, "Quantification of Clinical Material," the author discusses the frequent need for nonlinear statistical procedures. He defines abstractly the concept of "patterning" of variables for the benefit of statisticians as the nonvanishing of mixed partial derivatives of the predictive function. The definition can be given in terms of interactions. The system may be described as patterned of order k if all $(k+1)$ variable interactions vanish but at least one k variable interaction does not vanish. An unpatterned system is one for which all interactions vanish.

The Statistics of Bioassay: with special reference to the vitamins. *Chester I. Bliss.* Reprinted with additions from Vitamin Methods, Volume II (Paul György, ed.), New York: Academic Press, Inc., 1952. Pp. iii, 445–628. $3.50.

Jerome Cornfield, *National Institutes of Health*

The term "bioassay" refers to a method of measurement in which one estimates the amount of a substance present in a preparation on the basis of its biological effects. It is to be distinguished from a physical or chemical method of measurement solely by the fact that the measuring instrument used is a living organism and may consequently be subject to a considerable amount of variation. Even though alternative physical or chemical methods may exist, a doubt as to whether they are measuring that component responsible for the biological effect of interest makes the use of methods of biological assay often inescapable. As long as the major problems of bio-

logical research were qualitative, use of these methods was restricted to a small number of fields, but with the increasing emphasis of recent decades on quantitative problems, bioassay methods have some into their own. Some of the newer areas in which bioassay methods provide a useful and in some cases indispensable tool of measurement, in addition to the traditional one of biological standardization, are virology, enzyme kinetics, endocrinology, protein metabolism, and vitamin research.

The systematic substitution of statistical methods for rule of thumb procedures in the field of bioassay began in the thirties, although there had been previous scattered applications. The statistical problems involved are far from trivial. The central problem, from which all others stem is: *given* (*a*) that the administration of a known amount of substance, X, is followed by some quantitative, but nevertheless variable, response of magnitude Y, (*b*) that the average value of Y over all possible samples is functionally related to X, and (*c*) that the administration of a preparation containing an unknown amount of this substance results in a level of response Y', then, what is an appropriate estimate, X', of the unknown amount of the substance in this preparation and within what limits may the estimate be trusted? Rule of thumb procedures will take one only so far in a problem of this type, so that increasing numbers of research workers have felt the need of familiarizing themselves with the applicable statistical methodology.

The present volume, originally published as a chapter in Volume II of *Vitamin Methods*, edited by Paul György, is now one of several texts designed to meet the research worker's needs in this respect. Its author has been a distinguished contributor, both as a biologist and as statistician, to the theory and practice of bioassay for the past twenty years and is entirely conscious of the difficult problems of exposition involved. He has chosen to meet them by concentrating on the computational aspects of statistical methodology, and at the very outset warns his readers that, "it will not be possible to develop fully the biological logic of each procedure and its mathematical basis not at all." The reader who conscientiously works his way through all the numerical examples Bliss gives will absorb in the process a number of fundamental statistical ideas which will stand him in good stead in the practice of bioassay. Whether there are any easier ways of realizing this objective is a question on which almost every statistician has an opinion and almost none has any evidence.

This reviewer missed a discussion of the logic of bioassay. On first exposure to the bioassayer's argument one is inclined to dismiss it as a particularly gross form of the fallacy of the undistributed middle. At first blush it seems as if the argument is: (*a*) this substance, say insulin, causes a particular biological response, say a fall in blood sugar; (*b*) the unknown preparation in which I am interested also causes a fall in blood sugar; (*c*) it therefore contains insulin. The bioassayer is in fact more sophisticated than this. He knows that as a matter of pure logic his conclusion is correct only if there are no other substances in his unknown with direct or indirect effects on

blood sugar. If he seriously doubts the validity of this assumption, he will investigate the effect on blood sugar of those substances other than insulin that he might reasonably expect to be in his unknown. If one or more of the substances investigated has important effects he will attempt to remove it before assaying for insulin. He will become impatient, however, if the possible mischievous effects of other uninvestigated substances are pointed out. The abstract possibility of error does not interest him. In this respect the bioassayer is like any other experimental scientist. He is interested in detecting and eliminating sources of error that might reasonably be presumed to be present; he is not interested in errors that could conceivably be present, but in whose existence he has no reason but formal logic to believe.

Nevertheless the assumption that there are no other disturbing substances in the unknown may never be strictly true, and in some important cases (e.g., the assay of ACTH), can be grossly untrue. To appreciate truly the havoc this can and does create one need only imagine the present state of physics if the ordinary balance were sensitive not only to the mass of the object being weighed but to a series of extraneous characteristics as well, such as its shape, chemical composition, internal energy and color, and if, in addition to this, repeated measurements on the same object were 100 times as variable as they now are. The statistician may argue that such matters, while interesting and important, are no concern of his, and he may perhaps be right. But if these matters are not discussed in statistics texts, where is one to find them considered?

Bliss emphasizes the prevailing point of view that each assay should be self-contained, and in particular that the dose-response curve for the standard preparation should be determined anew in each assay. This reviewer would not endorse this position without qualification. Even when the dose-response curve for the standard is known to vary "significantly" from assay to assay it is by no means impossible that the use of a previous well-determined standard curve will give a more precise estimate of the potency of an unknown preparation and that the degree of precision can be validly estimated using previous data on variation from experiment to experiment. The decision to allocate half of one's experimental material to yet another determination of the dose-response curve for the standard should clearly depend on the magnitude and not the mere existence of variations in the level of this curve from experiment to experiment.

There are other aspects of the subject on which one is tempted to write either a concurring or dissenting opinion, but to do so would give an entirely unbalanced idea of the book. We simply enumerate therefore the following (*a*) although the Latin Square is a very useful experimental design in bioassay, methods of analysis that are appropriate to the argicultural field trial may not be appropriate in bioassay. In particular, it may not be necessary as it is in the agricultural field trial to select the Latin Square at random, and in consequence the appropriate error terms need not be the same; (*b*) Bliss' use of the same symbol, *lambda*, to stand for the ratio *st. dev.*/

slope in an assay with a continuous response and for the ratio *unity/slope* in an assay with all-or-none response is unfortunate, since if an assay with continuous response and one with all-or-none response are characterized by the same value of *lambda* as defined, then they will not have the same inherent precision; (*c*) it would be desirable to mention the existence of other methods of handling the all-or-none response then the probit method.

But to dwell any further on matters such as these would be misleading. J. H. Burn, writing in 1930, said that, "Biological assay, as carried out by the majority of workers in the world, still remains a subject of amusement or despair, rather than for satisfaction or self-respect." Twenty-five years later the situation has vastly improved and at least partly responsible for this are expositions such as the present in which instructions on computation procedures have been used as the vehicle for conveying statistical ideas. The book deserves a wide audience.

Demographic Yearbook, 1954. Sixth issue. United Nations, Statistical Office, New York: 1954, pp 729 $7.50 cloth, $6.00 paper.

KURT MAYER, *Brown University*

SINCE 1948 the Statistical Office of the United Nations has published a *Demographic Yearbook* containing regular series of comprehensive and systematic international population statistics. Because of the large and increasing amount of new data available each year, the editors decided at the outset to adopt two major modes of presentation Certain tables, regarded as basic or "standard," are repeated in each issue with appropriate revisions and the introduction of new information. Other subjects and special topics are rotated, being featured only in certain years Every volume contains a very useful cumulative subject index which makes it possible to determine quickly what subjects and time periods were included in specific issues of the Yearbook.

The basic tables which are repeated in every volume are concerned with censuses and estimates of the total population of over 200 geographic areas, with population density in these areas, with age and sex composition, and with basic vital statistics: births, deaths (including infant mortality), and marriages. Among the subject matter which is rotated are tables relating to marital status, economic characteristics, fertility, divorce, life tables, and international migration. In the 1949–50 volume the topics receiving special emphasis were marriage and fertility; in 1951 the spotlight was on mortality. The 1952 issue was devoted mainly to the geographical distribution of the population, including a special chapter on urbanization trends. The 1953 volume was limited to the basic demographic tables only, but the 1954 issue again presents a special emphasis, with natality the featured subject.

The 1954 volume is the most extensive issue published to date. It contains a total of 44 separate tables. 20 tables deal with statistics of live birth or stillbirth, and 8 of these are entirely new. The introductory and explanatory

apter "Technical Notes on the Statistical Tables" has also been expanded. every issue this section explains the scope and geographic coverage of the bles presented and provides a critical evaluation of the quality of the data th in a general sense, and specifically for each table. Like the preceding ues, this volume also contains a bibliography of recent official publications ntaining census data and other demographic statistics for each area of the rld. The work concludes with a country and subject index of the tables ntained in this volume, and the cumulative index mentioned earlier.

here is no doubt that the current issue, like its predecessors in this series, nstitutes a most valuable reference work both for the specialized scholar for the general reader interested in comparative international population tistics. But looking back over the six volumes which have so far appeared this series, it is perhaps timely to remind the editors of a hope, expressed in introduction to the first volume, of including some of the many topics itted from the first compilation by means of the rotation system. In te a few cases this hope has not yet been fulfilled: among the important ms which have remained consistently absent from the *Demographic Year-* are data on race, language, and religious affiliation, for which existing tistics seem adequate enough to warrant their inclusion. The present re- wer is also looking forward to a reappearance and extension of the tables taining economic and labor force statistics, not included in any of the 4 volumes. In general, a somewhat one-sided emphasis on purely demo- hic data at the expense of statistics pertaining to the social characteris- of the population seems to be the major shortcoming of this otherwise ellent series.

---

ulation Statistics and Their Compilation. *Hugh H. Wolfenden.* Chicago: The versity of Chicago Press, 1954. Pp. xxiii, 258. $7.50.

ANSLEY J. COALE, *Princeton University*

PULATION *Statistics and Their Compilation* was first published in 1925. The book here under review is the second edition. The original chapters e all been expanded and amended in response to developments since the edition, and two new sections (on forecasting mortality rates and on theory of reproductivity) have been added. There is also a new appendix W. E. Deming on "Some Theory in the Sampling of Human Populations." Volfenden is a distinguished actuary, and his book seems to have been ten primarily for actuaries. He says in his preface (p. viii) "... special ntion has been paid to the needs of actuarial and other students ...." it is clear that the actuarial rather than the other students were upper- in the author's mind. The central half of the book is occupied by the blem of the preparation of mortality tables. In most instances even non- uarial topics are treated with emphasis on points of special interest to aries. For example, in discussing "the reliability of census and registration istics, and the nature of errors therein" (section IV), the author restricts

his analysis to misstatements of age and omissions of certain age groups in censuses, and to defects in vital registration—the errors which present problems for the maker of mortality tables. Misstatements of school grade completed, occupation, income, race, etc., are not discussed. The half page devoted to discussing statistics of unemployment emphasizes their usefulness in calculating the costs of unemployment insurance—an actuarial problem.

Despite the emphasis on the actuarial point of view, the non-actuarial reader will find rewarding material in the book. The opening sections on population censuses and the registration of vital statistics contain useful summaries of the principal methods employed in various countries. One of the new chapters contains a succinct description of the most important concepts which have been developed for analyzing reproductivity. It also summarizes the limitations of these concepts. Not least, the non-actuarial statistician or demographer can get from this book, if he is willing to make the effort, an authoritative introduction to actuarial techniques and an understanding of the principal problems underlying the construction of mortality tables.

One of the features of *Population Statistics and Their Compilation* is the large number of references to demographic, statistical, and actuarial literature that it contains. Five to ten references per page is not unusual. Thus the reader is provided with a running bibliography for each topic. Unfortunately, the references are nowhere brought together in a systematic list. Some of the references occur in footnotes, most are set in parentheses in the text, and many are part of the text itself. An annotated bibliography arranged by subject would be a more usable guide to the literature. Deletion of the references from the body of the text would also make material which is difficult enough more readable. There is no index, but a detailed analytical table of contents.

Addendum to Demography, Developments during 1949–54. *Peter R. Cox*, Cambridge, England: Published for the Institute of Actuaries and the Faculty of Actuaries at the University Press, 1955. Pp. viii, 64, 5 s. net.

MORTIMER SPIEGELMAN, *Metropolitan Life Insurance Company*

THIS addendum brings up to date the content of the author's *Demography* which was published in 1950.* The new material is keyed to the paragraphs of the original volume so that it may be read in conjunction with it. Apparently a revised volume is in preparation.

The addendum contains succinct summaries of major British developments in demography within the last five years. These include, principally, the reports of the Royal Commission on Population and those that have so far been published on the 1951 Census of Great Britain. Together with the summaries are included comments that have been offered on these reports

* See review by Frank Lorimer in this *Journal*, Vol. 47 (1952), p. 349–50 June 1952.

by a number of critics. The summaries and comments, taken together, make a convenient reference.

A number of other topics are treated by the author, such as developments in vital registration in Great Britain, advances in morbidity statistics, measures of infant mortality, and a limited number of recent demographic data for Great Britain and the world.

Although intended essentially for actuarial students within the Commonwealth, the addendum and its parent volume provide, more generally, a valuable insight into British demographic thought and practice.

Cohort Fertility: Native White Women in the United States. *Pascal K. Whelpton.* Princeton: Princeton University Press, 1954. Pp. xxv, 492. $6.00.

CHARLES F. WESTOFF, *Princeton University*

THE major factors motivating the preparation of Whelpton's monograph are the analytical and interpretive problems created by the interruption of the long-time trends of the birth rate in the United States. It is well known that annual birth rates and the average size of completed families had been on a downward trend in this country for a long period of time, trends that were especially evident in the depression of the 'thirties. Many demographers, including Whelpton, developed population projections which involved the assumption of a continuing decline or stabilization of these low birth rates. In the light of the recovery of the birth rate in the 'forties, it was not surprising to see how far short these forecasts fell. Whelpton, among others, has been reluctant to accept the view that the sharp rise in annual birth rates implies a reversal of the trend toward smaller families. He has persistently maintained that one must apply sharper analytical techniques to birth data in order to ascertain whether the recent changes have not been due more to "temporary" changes in the time at which married couples decide to have their children, than to any basic alteration in the number of children ultimately reproduced. The application of cohort analysis, the calculation of the cumulative number of births occurring throughout their reproductive period to a group of women born at the same time, represents Whelpton's attempted solution of this problem.

The monograph accomplishes two primary objectives: the detailed description of a technique of fertility measurement (cohort analysis) and the presentation of over 100 pages of fertility tables for native-white cohorts born in each year from 1875 to 1933. The fertility rates included in the first tables are cumulative birth rates by order of birth by age of woman (Tables A, B). The next set of tables presents the distribution by parity of the women surviving to each age from each initial birth cohort (Tables C, D). The following two tables (E, F) present the calculations of the probability of bearing a child of n order during each age for the women of $n-1$ parity living at the beginning of each age. The series is concluded with two tables (HA, HC) which relate to hypothetical cohorts up to the year 1948.

These latter tables are constructed on the principle of life tables and reproduction rates.

The text is based largely on the rates presented in these basic tables. The author enumerates in elaborate detail the kinds of biases and errors from such sources as underenumeration and underregistration. The whole tone of the book, as a matter of fact, is of this nature. Whelpton's writing is characterized by extreme cautiousness and fidelity to detail.

In substantive terms, the author's historical analysis reveals the pattern of a recent making up of births postponed during the depression, an abnormally high rate for first births partly as a consequence of high marriage rates, and a long-time decline in family size as a result of sharp decreases in higher order births. His concluding chapter on the future course of fertility suggests that little change will occur in recent patterns of age at marriage. His medium forecast for the total birth rate for the cohorts of 1945–49 represents approximately the estimated rate for women of the 1905–09 cohort, although some intervening cohorts will probably at least temporarily interrupt the historical downward trend of completed family size. His forecasts of additions to cumulative birth rates from 1949 to 1963 indicate fewer first order and fewer high order births than in 1944–48, and more births particularly of third to sixth orders. The general forecast indicates a substantial increase of total births up to 1954 and a decrease thereafter. The forecasts prepared are not mathematical projections of past trends but empirical estimates of the most probable events based on the author's opinions of trends in age at marriage, attitudes toward family size, the treatment of sterility, and the like. His High, Medium, and Low categories are chiefly determined by explicit assumptions of differing future economic conditions.

From the point of view of contributions to statistical theory, nothing is claimed in this volume The contribution made to the methodology of population research is contained in the author's consciousness of the necessity of analytical refinement and the use of statistical controls, as well as in his critical appreciation of the need for actual measurement through generational time. It would be something of an understatement to say that the book is difficult reading. This derives not from any inherent complexity of the techniques described or the analysis presented, but rather from the technical nature of the subject itself and from the frequently cumbersome labels attached to the rates presented.

One general criticism that has been levelled at Whelpton's work deserves reiteration here. His attention to completed fertility occasionally results in the treatment of other fertility variables, such as annual birth rates, the timing of births, and so forth, as variables requiring control or as factors tending to obscure "real" trends. Although his conceptualization of such factors is appropriate to the analysis of completed fertility, the unfortunate picture results that other fertility variables are somehow less legitimate concerns, a point of view certainly not held by many social scientists and probably not intended by the author himself.

With respect to his forecasts, Whelpton voices the now standard defensive complaint that economists would have more justification in criticizing population projections if their own forecasts were more accurate. It would be more helpful at this juncture to declare a cessation to mutual recrimination and direct this energy toward collaborative effort.

There is no doubt that Whelpton has made a basic and significant contribution to the methodology of demography, both in terms of the research techniques developed and the construction of useful basic data tables.

Economic Problems of an Aging Population. Part 2 of Retirement Policies and the Railroad Retirement System. *Report of the Joint Committee on Railroad Retirement Legislation.* Washington, D. C.: United States Government Printing Office, 1953. Pp. xvii, 172. $0.45. Paper.

ERNEST W. BURGESS, *University of Chicago*

This report is issued by the Joint Committee on Railroad Retirement Legislation of the Senate and House of Representatives, Paul H. Douglas, chairman; Robert A. Wallace, staff director. Its purpose was to present a comprehensive survey of the economic problems of the aging population derived from over 100 sources by summarizing "existing statistical and other data regarding the economic problems caused by aging, generally, in employment and retirement." It was addressed not only to members of Congress but to private citizens and public agencies active in this field. It is factual and analytical and makes no recommendations.

A few main findings of this report may be briefly stated. The proportion of persons 65 years and over has greatly increased because of the increase in life expectancy and of the decrease in proportion of children and younger persons. The proportion of older persons in the labor force has steadily declined. Those still employed have small incomes in comparison with the total working population but more adequate than the incomes of the retired. Self-employed individuals constitute only 18 per cent of the entire labor force but 40 per cent of those 65 years of age and over. Older as compared with younger persons are at a disadvantage in having lower incomes, poorer housing, higher incidence of chronic and disabling illnesses, more need of medical care and less hospitalization insurance.

Findings of studies are presented showing that older workers compare equally and in some instances favorably with lower-age brackets in the frequency of disabling illness, in lower accident and absentee rates, less labor turnover. The results of a survey of over 3,000 companies by the National Association of Manufacturers in 1951 is quoted in which 99 per cent of companies reported workers 45 years and over as superior or equal to younger workers in work attitude and 93 per cent report them superior or equal in work performance.

The growing importance of retirement is indicated by the doubling between 1900 and 1940 of the number of years a man of 60 can expect to spend in retirement.

The report contains 75 tables and 33 charts. A valuable feature is a digest of 108 recent selected references on the problems of aging. It has an excellent index.

America's Resources of Specialized Talent. *Prepared by Dael Wolfle* New York: Harper and Bros., 1954. Pp. xvii, 332. $4.00.

ARTHUR H. BRAYFIELD, *Kansas State College*

THE wise utilization of human talent is a central problem in a democracy. Urgent national concerns must be balanced against the ideal of free individual choice of occupation. Special attention to the gifted may encounter public distrust of an intellectual elite.

The appropriate starting point for any constructive efforts is careful appraisal based on adequate factual data. This the Commission on Human Resources and Advanced Training has attempted to provide in a long-awaited report by Dael Wolfle. The Commission, an outgrowth of the experience with manpower problems of four national research councils in World War II, spent nearly four years and approximately a quarter-million-dollar Rockefeller Foundation grant in the development and analysis of relevant information.

This volume summarizes a comprehensive survey of the specialized manpower resources and requirements of the United States The Commission was concerned with "those persons who are educated, intelligent, able to work with ideas, and qualified to plan and understand and direct the nation's complex web of industrial, technological, social, scientific, and governmental institutions and problems." Its studies were organized around three interrelated problems. First, what is the present supply and what is it like? Second, what are the demands now and what are they likely to be in the future? Third, what is the potential supply?

Among the methodological problems confronting the Commission staff at the outset were the necessities for (1) a sharpened definition of specialized personnel; (2) a classification of fields of specialization; (3) criteria for determining membership in a field, and, closely related; (4) criteria for determining the allocation of jobs to fields. These problems were resolved in a fairly satisfactory manner and represent a substantial contribution to manpower studies.

Where possible, the Commission used readily available data. In addition it conducted, or contracted for, first hand investigations. Although an appendix particularizes some of the methods of data collection and analysis it is not possible to evaluate the procedures adequately. The Commission was aware of limitations and Wolfle candidly says that "one could well ask for improvements in practically every type of statistical finding and estimate made in this book." But the figures "are offered as the Commission's best estimates for the United States as a whole in the middle of the twentieth century." And the fact that earlier fragmentary studies bearing on some of these same issues support many of the Commission's conclusions gives one considerable confidence in the major outlines drawn in this report.

A detailed account of the findings is obviously impossible. The following gives something of the nature of the enterprise.

The proportion of college graduates in the appropriate age range has steadily increased and may be expected to continue. The composition of the college graduate population is constantly changing. Smaller proportions come from the humanities and the natural sciences. The rapid increases have come in education and business administration which now account for nearly a third of all bachelor's degrees! The social sciences show a fairly steady relative increase.

About 2 per cent of bachelor's degree recipients go on to the doctorate; the percentage is slowly increasing. The distribution of doctor's degrees among fields of specialization has been comparatively stable, with the natural sciences accounting for the largest percentage of degrees granted. Wolfle suggests that the character of the degree may change with "applied" degrees increasing.

Continuance in the field of specialization comes in for extensive consideration. About one graduate student in three shifts his specialization from his undergraduate major. Approximately 38 per cent of employed college graduates work in the field of their undergraduate major. Shifts are most likely to occur among liberal arts and sciences majors. Contary to the frequent interpretation of such findings as evidence of the inadequate utilization of human resources, the Commission points to a positive social value in the form of desirable flexibility in the labor supply. It is interesting to observe that economists have been making a similar virtue of labor mobility. These are prosperous times.

Short-run supply and demand projections made by individual areas of specialization were based on a variety of sources of information of apparently uneven value. Although these projections make interesting reading (unfortunately statisticians were lumped in with other groups), the attempt highlights the very great need for adequate data.

It was not thought feasible to estimate the long-range potential supply for individual fields. Instead, an analysis was made of the potential for college attendance and the factors affecting attendance and graduation. This portion of the report is fairly conventional but does bring the documentation up to date. A major conclusion is that a lot of bright youngsters do not get to college, and the important reason seems to be a lack of adequate motivation.

Also documented anew is the frequently observed fact that "students in some fields are more highly selected in terms of intelligence test scores than are students in other fields." (Could the great increase in education and business administration enrolments be related in part to the finding that these students are well toward the bottom of the ability hierarchy?) The report neglects the comparatively large literature dealing with the *differential interest patterns* of students entering and continuing in specialized fields. It also minimizes the practical use of differential aptitude measurement to an extent that disturbs me.

The discussion of methods for improving the utilization of the potential supply covers familiar ground but has the virtue of bringing various proposals together in a systematic appraisal of their possible effectiveness. It may be news to some that scholarships and financial aid are of limited value.

Although the amassing of comprehensive data of the type presented here is a very worthwhile contribution, particularly as it focuses national attention on the problem, the reviewer felt "let down" after completing the book. A half dozen adequately designed studies directed to specific aspects of the problem might have contributed really new knowledge. For example, recent advances in the measurement of motivation for achievement make it possible to test empirically hypotheses regarding the influence of different variables upon such motivation. Controlled experiments with college attendance as the dependent variable are feasible. Support of a really adequate venture in differential prediction with, say, the Yale test battery, has potentialities. The commission missed a fine opportunity to outline, at least, needed research.

In summary, the Commission's report has general information of real value. It demonstrates the difficulties in collecting relevant data and forcefully reminds us of the need for continuing, systematic appraisal. The recommendations for improved utilization of specialized manpower make sense. An emphasis on freedom of occupational choice permeates the entire report and is impressive evidence of the faith of the Commission and staff in the validity of the democratic process.

The Uneducated. *Eli Ginzberg and Douglas W Bray.* New York: Columbia University Press, 1953. Pp. xxv, 246. $4.50

Z. W. Birnbaum, *University of Washington*

During World War II, more than 700,000 men were rejected by the Armed Forces on the grounds that they were "mentally deficient," and an additional 300,000 were rejected on the same grounds in the year following the outbreak of fighting in Korea These startling figures brought into focus the basic problem treated in this book· the waste of human resources, possibly due to the fact that a sizeable fraction of the population appears unable to perform mental tasks which are supposed to be taught in the first few years of elementary schooling.

In the first part of the book entitled "Education and Society" an attempt is made to trace the history of this defect in the development of elementary human resources in the United States. Narrowing it down, somewhat arbitrarily, to "illiteracy," the authors make use of census data between 1870 and 1950 to uncover the trends within various social groups and geographic areas of the United States. Particularly striking differences are observed between the native born white and the Negro population, and between the states in the Northeast and Far West on the one side and those in the Southeast on the other. The rejection data of World War II are then broken down

and analyzed in a similar manner according to geographical regions and two main groups: white and Negro. Remarkable relationships are observed between rejection rates and educational expenditures per pupil, per capita income, and percentage of rural population in the different states.

The second part of the book, under the heading "Military and Civilian Performance," deals with the poorly educated in the Armed Forces during World War II. The mobilization plans, prepared against a background of the large scale unemployment of the 1930's, did not contemplate the possibility of a shortage in the nation's manpower resources, but aimed at securing for the Armed Services the better men out of the total pool. In the beginning of the war the Army inducted only men who had "the capacity of reading and writing the English language as commonly prescribed for the fourth grade in grammar school." With the steady increase of manpower shortage, however, the Armed Forces had to relax these admission standards. From 1943 on, a psychological test was substituted for the criterion of four years of school or equivalent training, and those who passed the test, even if illiterate, were inducted and assigned to Special Training Units in which they were to be educated to a level corresponding to at least the fourth grade.

A major effort has been made by the authors to study the effectiveness of these Special Training Units, by exploring the civilian background of soldiers assigned to such units and their adjustment to civilian life, comparing those who were graduated with those who failed to graduate from the STU's, and appraising the performance of the graduates in subsequent military service. The approach to this group of problems is methodically sound: a sample of individuals assigned to STU's was obtained, the records for each of them were traced down to a remarkable detail, and these records were used as material for a statistical analysis. The statistical design for the collection of data, however, could have been improved. The soldiers in the STU's were classified into eight groups according to the categories: Northern, Southern; White, Negro; year 1943, year 1944. A sample of 50 individual records was then obtained from each of the eight groups (by a procedure which is not described in detail). While the over-all sample size of 400 may be large enough, the allocation of equal quotas of 50 to the eight strata regardless of their size can hardly be considered efficient statistical procedure.

One of the salient conclusions reached by the authors in the study of the 400 individuals is that this substandard human material was to a very great extent rehabilitated by the STU's and re-made into good soldiers. A comparison of STU graduates with a control group consisting of soldiers who did not require remedial training suggests that the STU graduates performed less well than those in the control group but nevertheless the STU's represented a clear gain to the Army.

The study dealing with the Special Training Units was rounded off by following up the graduates of these units after their return to civilian life.

The information obtained, although fairly incomplete, seems to indicate that a majority of the graduates considered the training received in the STU's as helpful in their lives.

In view of their finding that the main concentration of uneducated is in the South, the authors undertook a study aimed at finding out to what extent the problem of those lacking the most elementary education affects the supply of labor for industry in the South. The evidence gathered appears to present the following picture. the educational facilities are still very uneven, and the poorly educated and the uneducated have found considerable difficulties in making a place for themselves in southern industry; industrial management in the South, however, is generally satisfied with the educational background of its working force, due to the fact that southern industry, although growing rapidly in recent times, has lagged behind the expansion of southern education and has encountered no difficulty in obtaining all the workers with the required educational background it needed.

A separate chapter entitled "The Uneducated Migrant" is devoted to the problem of the uneducated among those who, mainly due to economic pressure, change their domicile and try to make their living elsewhere

In the third part, under the heading "Human Resources Policy," the authors attempt to interpret the results of their study in terms of implications for public policy. After presenting considerable material on the development of schools in the geographic area which they have found to be most critical, the South, and carrying out comparisons between the South and the rest of the country, they utilize the information so assembled to discuss the "limited question of whether the white and the Negro children born in the South during the 1940's and 1950's, including the more impoverished rural areas, now have an opportunity to acquire a basic minimum education." The authors admit that the evidence is not clear cut. They point out that there has been an increased effort made by the southern states to close the gap between their rural and urban schools and particularly between Negro and white schools, and that the southern states have been raising their educational standard rapidly. In spite of these striking improvements, some numerical evidence and opinions expressed by various state departments of education lead the authors to the statement that "the fact must be faced that a considerable, if unknown, number of children, particularly in the rural south, and more particularly among the Negroes, will fail to complete four grades of schooling and acquire basic literacy."

In a chapter "The Armed Services" the present practice of the Armed Forces of rejecting those individuals who fail to pass the Armed Forces Qualifications Test (AFQT) is subjected to a critical discussion. Among other things, it is pointed out that by this procedure the Armed Forces may be losing a substantial number of potentially good soldiers who could be reclaimed by a specialized training program.

In a concluding chapter, "Human Resources Potential," the authors summarize their findings with the aim of formulating desirable actions for

the future. Having localised the focus of the problem in the southern states and, to a very much smaller extent, in such groups as migrant workers, they reach the conclusion that an effective remedy would consist in Federal action directed towards raising the level of education in the economically less well-to-do states. This, in their opinion, would attack the evil at the source, while such steps as admitting poorly educated individuals to the Armed Forces and establishing remedial training for them would help reduce its effects.

The book makes interesting reading and the conclusions are in essence well substantiated. One can not help, however, thinking that the quality of the presentation and the validity of the findings might have been considerably better had this large scale investigation utilised some modern statistical techniques. Although the book relies heavily on statistical evidence, the statistical methods used do not go beyond tabulating percentages. The word "significant" is used to mean that some figures appeared to be significant to the writer. As already mentioned, the planning of the study of the Special Training Units and the evaluation of its results could have been improved. There also are quite a few slip-ups, such as 11.5 per cent being referred to as "one in ten" (p. 21), or a reference to a non-existing table (p. 31, second paragraph), or statement "The reader will recall that 38 graduates . . . were so classified" (p. 110) which made this reviewer search for the number 38 throughout the book.

In spite of minor shortcomings, *The Uneducated* is a thought-provoking book. It formulates and discusses a real and important problem and points out possibilities of attacking it. It certainly constitutes a constructive approach to the area of conservation of human resources and makes the reader very much aware of the need for much further work in that direction.

Towards an Understanding of Juvenile Delinquency. *Bernard Lander*. New York: Columbia University Press, 1954. Pp. xv, 143. $3.00.

C. C. VanVechten, *U. S. Navy Bureau of Ordnance*

The basic materials of this study are 8,464 official records of presumably different individuals before the Baltimore Juvenile Court over the four year period 1939–42, and the 1940 census tract data for the city.

The problem of definition is recognized. Lander uses the term "juvenile delinquent" to mean "brought into court on a delinquency petition." He has no data on the characteristics of census tracts which influence the probability that specific acts will result in formal hearings and does not speculate on them.

Statistical techniques include zero order, partial and multiple correlations. A chapter is devoted to factor analysis. Parabolic functions are fitted to data that appear non-linear. Significance levels are meticulously stated.

In addition to delinquency, the factors used are medians for years of school completed and monthly rent and percentages for overcrowding, home owner-occupied, home substandard, non-white, and foreign born.

Lander is primarily interested in examining the Shaw-Burgess concepts of the causation and ecological distribution of juvenile delinquency. This reviewer finds it possible to agree with many of the concepts and conclusions without agreeing that they contravert the Shaw-Burgess idealized ecological models. The difference seems to center on whether the "Chicago School" ecological model is taken as a rigid statement of what "should" uniformly exist or as a useful scientific fiction to aid in thinking and valid in its details only when other factors are constant. Lander seems to make the former assumption and feels he has disproved the Shaw-Burgess contentions when he can demonstrate specific exceptions even though the exceptions are just what any ecology student would predict given all the facts.

For instance it is hardly startling that there are high rates contiguous to an industrial area incorporated into the city 256 years after its founding. Nor is it startling to find that the role of the immigrant has changed now that he has been displaced by the Negro as newcomer and low man on the economic totem pole. One highly interesting fact, new to this reviewer, is the demonstration of a marked curvilinear relationship between delinquency and percent Negro with the rate maximum where the ratio is about 50–50. Old areas of stable Negro occupancy showed up with very low rates despite extremely poor status on the economic indexes.

One learns a great deal about Baltimore although it is regrettable that with 13 maps of the city none identifies the tracts so constantly referred to in the text.

Lander finds the delinquency rate fundamentally related to *anomie* (which term he prefers to "social disorganization" as indicating isolation from the pressures for conformity within the dominant culture) with socio-economic factors in a distinctly secondary role.

The distinction between cause and covariance is carefully maintained. This reviewer has some questions about some of the statistical techniques; perhaps because it is not always clear just what happened. (Are all recidivists counted only in the first year, and only at the original residence? Why fit parabolic curves rather than use exponential transformations?) Particular exception must be taken to the contention that factors having zero order correlations of .69 and .73 have "no substantive relationship". . . (with). . . delinquency" because they have essentially zero partial correlations. This involves assumptions of orthogonality and normality in the rest of the data that are at best dubious and, as applied to the group of factors, would justify the conclusion that the sum of the contributions of the factors studied accounted for less than 10% of the variability in delinquency.

It is regrettable that typographical errors are generously sprinkled throughout the book starting on page one. The reader sometimes, as in Table XVIII, finds it difficult to supply corrections.

Despite limitations this book will be a "must" for anyone contemplating the use of census tract data to establish basic relationships between ecological and sociological phenomena.

Occupational Mobility in the United States, 1930–1960. A. J. *Jaffe* and R. O. *Carleton*. New York: King's Crown Press, 1954. Pp. xiv, 113. $2.75. Paper.

HERBERT GOLDHAMER, *The RAND Corporation*

THE primary aim of this study is to provide estimates of the size of the United States male working force and its distribution among ten major occupational groups for the year 1960. Four projections are provided, two on the assumption of depression conditions similar to the 'thirties and two on the assumption of prosperity conditions similar to the 'forties. The two projections of each of these pairs are distinquished by their incorporation of different assumptions concerning the number of men in the armed forces.

Utilizing the 1930, 1940, and 1950 censuses of occupation the authors trace the occupational composition of successive age cohorts through these time points, each age cohort then being projected to 1960; the major occupational groups are the summation of the age cohort projections. The changes in each age cohort through time are derived from estimates of four components: deaths, retirements, new entries, and net mobility (inter-occupational shifts). These component projections lead to final estimates rather different from and doubtless more accurate than those that might be secured from simple extrapolations of past census occupational distributions.

The civilian male 1960 working force is estimated to range from a minimum of 45.0 to a maximum of 46.8 million (as compared with 42.3 million in 1950). The estimates of working force distribution among the ten major occupational groups are relatively insensitive to the two different assumptions on the size of the armed forces. The "depression" and "prosperity" assumptions are introduced by assuming for one projection the retirement, new entry, and net mobility rates of the 'thirties and for the other projection those of the 'forties. Professional, technical, and kindred workers show a percentage increase in 1960 (over 1950) under both sets of assumptions; farmers and farm managers a decrease in both cases; and sales workers and operatives virtually no change in both cases. Occupations showing an increase under depression assumptions and a decrease under prosperity assumptions are: service workers, farm laborers and foremen, and laborers except farm and mine. Occupations showing a decrease under depression assumptions and an increase under prosperity assumptions are: managers, officials and proprietors (except farm), and craftsmen, foremen, and kindred workers.

To provide these estimates the authors had to reduce the 1930, 1940, and 1950 censuses of occupation to a common occupational classification, a task which they have performed with exemplary care, although their success was necessarily qualified at some points by the limitations of the census data. A considerable part of the volume provides analyses of the separate components that went into the final working force distribution estimates. The component data, however, while constituting a proper basis for the final working force estimates have relatively little value in themselves. This is

particularly the case with the data on net occupational mobility which turn out, after closer inspection of the methodological section, to reveal little of interest on inter-occupational changes. The authors have, however, taken material from Gladys Palmer's *Labor Mobility in Six Cities* (New York: Social Science Research Council, 1954) which contains data on the occupational changes of specific individuals and reworked these into an interesting table (Table 13) which shows men classified by major occupational group both at the beginning (ages 15–24) and end (ages 55–64) of their working careers. This table would have been of still greater interest had the initial age group been somewhat older. Among the 15–24 age group are many young persons whose initial jobs are of a transient character. To compare these with the jobs they hold at 55–64 is not of much interest. Nonetheless this table is decidedly useful and does more to justify the title of the work than the authors' lengthier discussions of net mobility.

Immigration and the Foreign Born. *Simon Kuznets and Ernest Rubin.* New York: National Bureau of Economic Research, Inc., Occasional Paper 46, 1954 Pp. xii, 107. $1.50.

HOWARD G. BRUNSMAN, *Bureau of the Census*

PART I of this brief report contains an analysis of the basic trends in the foreign-born white population of the United States and in immigration into the United States in the period since 1820. The contribution of the foreign born to total population and to the labor force, and the cyclical variation in immigration and the relationship of these variations to business cycles are considered.

Part II of the report presents a summary statement of the methods used in preparing estimates of the foreign-born white population of the United States based on the migration-survival method by decades since 1880. These estimates are presented by sex and age groups and are compared with the statistics obtained in the decennial census. Estimates of number of foreign-born white in the labor force are also presented. The fluctuations in these various series are analyzed in greater detail than in part I.

Part III of the report contains a thoroughly-documented statement of the statistical methods followed in evaluating the series, in preparing the estimates, and the problems faced in this procedure.

It is stated in this report that "Restriction of immigration had at least one result that may have decreased the stability of the economy. . . . Swings in immigration and in the rate of increase of native-born and native parents are roughly synchronous. . . . But swings in rate of increase of native-born of foreign parents lag a decade behind immigration. . . . Because of this difference in phase, the net additions to population from the two sources together are steadied." An examination of table 9 of the report shows greater stability in the native-born of native parentage than in total population. This indicates that the damping effect of the lag in births to the foreign

born is not sufficient to offset the greater range of fluctuation in immigration.

The estimates of the number of foreign-born white show basically the same level and fluctuations as the figures obtained in the population census. This similarity is especially gratifying in view of the quite serious limitations of the basic data, including the limitations of the census counts for the earlier years.

It is refreshing to examine a brief report with a thoroughly-documented statement of the methods followed in preparing the statistical series which are presented. Most analysts are impatient to interpret the results of the studies and devote but little space to the explanation of how the basic series were developed. Progress in the analysis of data is advanced when a student is able to build on the work of his predecessors. This cumulative approach is feasible only when the student is able to review and evaluate the statistical basis of the earlier work. Part III of this report may be commended as an excellent example of the proper presentation of the methods and limitations of the basic underlying data.

This is an interesting and workmanlike job and one must agree with the authors that "the area of investigation touched by this paper has long been slighted and might richly repay more intensive research."

British Post-War Migration. *Julius Isaac.* National Institute of Economic and Social Research, Occasional Paper XVII. Cambridge: Cambridge University Press, 1954. Pp. xix, 310. \$6.00; 30s.

WILLIAM PETERSEN, *University of California, Berkeley*

As ALL who have used them know, international migration statistics are not very satisfactory. The definition of the area is vague, since the arbitrary distinction between migrants and visitors leaves many ambiguous cases, the count is typically neither complete nor accurate; the classification is usually into categories that are more or less irrelevant to an analysis of population trends.

Together with an anonymous "post-graduate research worker," Isaac collated the information given on shipping manifests for a ten per cent sample of all British emigrants during 1946–1949. This independent set of data made it possible for him to discuss intelligently matters not included in the Board of Trade passenger returns—for example, the emigrants' family structure and the region in Britain of their prior residence. With these data, he also subjected the standard statistics to a thorough critical review, and by the time the volume was published some of his suggested improvements had already been adopted. This section of the volume, roughly the first third, is at once the most original and the most stimulating, and well worth the close attention of anyone concerned with the analysis of migration trends.

There follow five chapters on the countries of future residence, with particularly detailed discussions of Australia and Canada; three chapters on

immigration to Britain; and two summary chapters, one on policy and one on statistics.

The discussion within this framework offers much that is of broader significance. The book begins, for example, with an attempt to distinguish in general terms between permanent and quasi-permanent, between recurrent and non-recurrent, and between visible and invisible migration. These are terms that might well become standard. The propensity to emigrate, to take another example, varied widely from one region of the United Kingdom to another, but it was not associated especially with regional differences in either occupational structure or population growth (p. 76). These negative findings challenge two widely accepted hypotheses on why people migrate, and Isaac discusses briefly the possibility of supplementing such data with public opinion polls (pp. 248 ff.).

According to Isaac's estimate, while some 1,084,000 persons emigrated from the United Kingdom during 1946–1950, net emigration amounted to only 140,000, or substantially less than an insignificant one per thousand population per year. Even so, the movement has been important, Britain has lost skilled laborers and technicians, clerical workers, businessmen and professionals; and has gained an almost equal number of unskilled and semi-skilled workers. In demographic terms, there has also been a loss, especially by the emigration of about 100,000 war brides and their 50,000 children. On the other hand, contrary to what one might expect, old-age dependents were about three times more frequent among emigrants than among British population (p. 212).

These migration trends have been the consequence, at least in part, of deliberate policy. The British government has encouraged and financed both emigration to the Dominions and immigration of aliens. One might ask whether this has been a reasonable process, or whether "Britain in the name of Commonwealth solidarity and strength has been making an unnecessarily heavy sacrifice of badly needed skilled workers" (*Economist*, August 28, 1954). Isaac is inclined to accept at face value the claim that subsidized emigration strengthens Commonwealth bonds, although it is at least possible that Britain's "half-hearted and inconsistent" emigration policy (p. 242) has had, on balance, the contrary effect For Britain's interest has been essentially opposed to that of the Dominions· "without the operation of the agreements the number of emigrants leaving essential occupations would have been substantially larger" (p 243). Isaac treads softly through this thorny area, and his discussion does not quite face up to the paradox of a government paying well for the loss of needed skills. More fundamentally, it seems to me that he has not yet resolved a dilemma evident in his earlier *Economics of Migration*. On the one hand, he patently favors free migration, both for its humanitarian value and for its relation to economic liberalism; on the other hand, he welcomes the state control of migration even though this is typically set by nationalist, not to say mercantilist, motives.

Apart from such highly debatable questions, this is an excellent monograph. While narrower in ostensible subject matter than the *Economics of Migration*, it should not be restricted to those especially interested in postwar British migration. Demographers generally have come to expect a high level and a wide range of genuine competence from Isaac, and they will not be disappointed in this work.

Year Book of Labour Statistics, 1954. *International Labour Office, Geneva, 1954.* Pp. xv, 397. $5.00.

JACK CHERNICK, *Rutgers University*

In this volume the I.L.O. continues its useful compilation of statistics on labor force composition and the economic position of the working population in countries all over the world. The data, derived from communications to the Office or from official publications of the countries covered, are presented in eleven chapters and an appendix. The scope of the volume is indicated by the chapter headings: Total and economically active population; Employment; Unemployment; Hours of work; Wages and labour income; Consumer price indices and retail prices; Family living studies; Social security; Industrial injuries; Industrial disputes; and Migration. Each chapter is introduced by brief notes (in English, French and Spanish) explaining the tables included. But the textual material includes no discussion of the original statistical compilations in the countries covered. For such discussion of definitions, methods of collection and classifications of data, the reader is referred to other publications of the I.L.O. and the United Nations.

Presumably, the collection in one volume of statistics on a world-wide basis is designed to serve two purposes: first, to make available in a central source the statistics on employment and labor conditions in a large number of countries; second, to permit international comparisons. The first purpose is quite adequately achieved in the *Year Book*, even though, in some cases, interest in a specific country or area of the world would, of necessity, have to be pursued beyond the data reproduced here. Useful aids to such research are included in this volume in the form of an index of countries included in each table (pp. 394–397) and a list of publications constituting principal sources of current statistics for the several countries.

Unfortunately, the second purpose often is not realized. This, of course, is in no sense the fault of the compilers, but is due to the inherent lack of comparability in many of the statistics prepared by national governments; there are wide differences in definition, techniques of collection, coverage, etc. However, the authors have managed, in some cases, to arrange the data so as to permit reasonable comparisons in certain narrow categories even where aggregate comparisons are not possible. They have, moreover, in the prefatory notes to each table, done a useful service in cautioning against drawing international comparisons whenever these are rendered invalid by the original character of the data.

It should be added that East-West comparisons are not possible for a more cogent reason than lack of uniformity in the statistical base; of the communist countries, figures are shown only for Czechoslovakia and Poland and these appear in only a scattered few of the tables.

Employment and Wages in the United States. *W. A. Woytinsky and Associates.* **New York: Twentieth Century Fund. 1953. Pp. xxxii, 777. $7.50.**

N. Arnold Tolles, *Cornell University*

This is the most monumental work on labor problems to appear since Millis and Montgomery's three-volume *Economics of Labor* (1938, 1943). In spite of its more specialized title, *Employment and Wages in the United States* covers almost as broad an area as the earlier work. Moreover, the magazine format permits Woytinsky to provide almost as much printed matter in a single volume as did Millis and Montgomery in their three volumes.

Millis steadfastly insisted that *Economics of Labor* was "not a textbook." The present volume has the characteristics of an encyclopedia to an even greater degree. Yet instructors have already found, in this case as in the earlier one, that selections from such sources often provide more effective teaching materials than do the standard textbooks. There are disadvantages, to be sure. Students will be reluctant to purchase or to carry a volume of this size. More important, the very definitiveness of the materials is apt to be stultifying, unless the instructor points up the issues, propounds challenging questions, and bridges the gap between the students' minds and the book. Woytinsky's work is written in a much more lucid style, but it is less provocative, than was Millis and Montgomery's.

As a reference, the present volume is superb, once the teacher or research worker has defined his own problem. The great advance in American labor statistics during the past 20 years is nicely demonstrated, especially in the areas of labor force, workers' earnings in relation to national income and wage structures. Millis and Montgomery (and their colleague, Paul H. Douglas) were adept in squeezing the utmost meaning out of the quantitative materials of their time. Woytinsky and his associates have mined with great skill the much richer ores of the present day, especially those of the Social Security Board, the Bureau of Labor Statistics and the Census Bureau. Both within the articles and in numerous appendixes, these authors have provided large masses of refined statistics which will serve as a starting point of original contributions for many years to come.

The selection of wages and employment as the fulcrum for the analysis of industrial and labor relations will seem old-fashioned to many persons, but this approach is heartily welcomed by the present reviewer. Indeed, his own complaint is that this focus is not kept in view with sufficient single-mindedness. In deference to current tastes, almost one-third of the work is concerned with "The Institutional Setting" (Part II). Very valuable information is presented therein, which helps to make the book an encyclo-

pedia, but the relevance of this material to wages and employment is never made clear.

Woytinsky has evidently sought to glean the most from a combination of theory, institutional history, and statistics. Personal evaluations of the result will be based on the judgment of each scholar as to which of these roads to understanding he considers to be most crucial. Admiration has been expressed for the handling of the statistics. The institutional history, however important by itself, appears to this reviewer to be in the nature of "filler" in a work of this kind.

As for theory, one notices at once the initial effort to relate the whole work to a theoretical framework. The very first chapter usefully recapitulates the nineteenth century wage theories, in a way that has a greater breadth but less penetration than the notable Volume I, Chapter IV of Millis and Montgomery. As for Woytinsky's first chapter, one is most puzzled by the concluding assertion that the "great names—from Adam Smith and Nassau Senior to Marshall and Pigou—can be called as witnesses in support of. . . . four principles," including the principle that "labor unions . . . are the mainspring that moves the mechanism of supply and demand in wage determination." The classical economists were certainly reformers but they were not to that extent convinced of the central role of labor unions. Indeed, much of the factual analysis in this book seems to put labor union efforts in their appropriate place, rather than to make them appear as the "mainspring" of the economy.

On its theoretical side, one misses any statement of a systematic theory of employment. While Part I (Wages) is introduced by an essay on theory, Part III (Employment and Unemployment) has no such introduction. It would seem that neither the influence of Keynes and his followers nor that of his critics and opponents have penetrated very deeply into the thinking of American labor economists.

The intended climax to this volume is the text of the Report of the Committee on Wages and Employment of the sponsoring Twentieth Century Fund. This harmless but ineffectual little sermon exhorts the parties to collective bargaining to be tolerant of each other and especially exhorts labor unions to moderate their wage demands so as to avoid the danger of inflation. If anyone wishes to know the kind of compromise statement which could be agreed to in the year 1953 by a cross-section of union, management, and public representatives, the Report may be an interesting historical document. At least, the inclusion of the Report is less annoying than that of a sponsor's TV "commercial." However, there is little evidence that the thinking of the Committee was influenced by the scholarly efforts of Woytinsky and his associates.

Some mention should be made of these "associates." There were sixteen of them. Their names are buried in one of the four title pages, and in the appropriate chapter-title footnotes, but they are not displayed in the otherwise excellent table of contents. While Woytinsky clearly organized and inte-

grated the whole work and indeed wrote about half of all the forty-four essays (averaging 12 large pages per essay), the "associates" would seem to deserve greater prominence. Like Woytinsky himself, all of the associates are, or recently have been, Washington civil servants and they deserve a recognition which our society seldom gives to such important contributors to our basic knowledge. The least this reviewer can do is to list their surnames, as follows: Arnow, Ball, Bowden, de Schweinitz, Douty, Ducoff, Evans, Grunfel, Huber, Kovner, Livernash, Ober, Pancoast, Sanders, Siegel, and Stocking.

The Influence of Plant Size on Industrial Relations. *Sherrill Cleland.* Princeton: Princeton University, 1955 Pp. 65. $2.00. Paper.

OSWALD HALL, *McGill University*

THIS is a study of the relations between plant size and industrial relations in eighty-two plants in Trenton, New Jersey. Plant size is measured in terms of numbers of workers, and industrial relations mainly in terms of the degree of conflict between workers and management The study bristles with the difficulties involved in trying to secure satisfactory numerical values for such things as income of workers, number of conflict situations, and so on. The actual correlations between plant size and pattern of industrial relations are not of a very high order.

The main variation between plants seems to lie between those which have retained fairly intimate, personal relations between managers and workers and those in which professional public relations personnel and specialized union functionaries stand between the two. These newly emerging semi-professionals, brought in to offset some of the effects of increasing size of plant, frequently bring a host of new problems in their wake. Hence it is not only plant size which influences industrial relations, but also the sorts of administrative machinery which have been set up to offset the disadvantages of size.

An International Comparison of National Products and the Purchasing Power of Currencies. A Study of the United States, the United Kingdom, France, Germany and Italy. *Milton Gilbert and Irving B Kravis* Paris: Organization for European Economic Co-operation, 1954 Pp 203. $3 00

JOHN H. ADLER, *International Bank for Reconstruction and Development*

THIS is an ambitious statistical study which attempts to do just what its title indicates: to compare the national products of the United States, the United Kingdom, France, Germany, and Italy, and the purchasing power of their currencies. The attempt has been eminently successful, the conceptual and statistical limitations of the undertaking notwithstanding

The objective of the study is the revaluation on a uniform price basis of the goods and services produced in 1950 (and, to some extent, in 1952), in each of the five countries. This required the gathering and sorting of in-

formation on the quantities of the various goods and services produced, and presented a pricing problem complicated by quality differences from country to country of the same kinds of goods.

From this exercise there can be derived five sets of figures showing the value of national products of the five countries at U.K., U.S., French, German, and Italian prices, and the purchasing power equivalents of the five national currencies. Actually, what the authors have done is to value the European national products at U.S. prices (and, to value the American GNP at the respective European prices), to compare the GNP's of the four European countries and the U.S. at an average European price level, and to compute purchasing power equivalents on the same bases. To obtain meaningful and comparable quantity and price data required the rearranging of the national accounts on a uniform—*or nearly uniform*—basis and a breakdown of the components of GNP into numerous subdivisions. (The main tables show a breakdown under 39 headings, but the text and subsidiary tables indicate, in many instances, a further breakdown.)

The results are striking, but by and large, in accordance with expectations, based on casual and piecemeal observations. If the national products are revalued on a uniform basis, whether at U.S. or European prices, the differences in per capita GNP between the U.S. and the European countries are very much less than if national product estimates are converted at prevailing exchange rates. To give but one example: average per capita consumption of the four European countries comes to U.S. $369, *if* valued at prevailing exchange rates, and to $615, if revalued at U.S. prices, compared with U.S. per capita consumption of $1,259. There are, of course, wide variations among the European countries: per capita GNP, at U.S. prices, is $829 for the U.K., $663 for France, $523 for Germany, and $388 for Italy. In every instance, the conversion of national values at prevailing exchange rates results in a much lower figure.

The detailed breakdown of the various components of GNP, the computation of purchasing power equivalents, and a series of other tabulations of results throws new light on a large number of old questions. For instance, it is most illuminating (and amusing) to see on what the average American consumer spends the $430 at his disposal over and above the $829 available to his British cousin. The American consumer eats a bit more (48 dollars worth) and, on the whole, better food: less cereals, potatoes, and fats and oils, but much more fruits, vegetables, and dairy products. He has a few extra drinks ($3), and he smokes a great deal more ($5). But the big differences come in five expenditure groups: purchase of household goods (furniture, equipment and appliances, household and personal supplies) on which the American consumer spends $82 more than the British consumer or nearly three times as much; purchase of transportation equipment (cars, motorcycles and bicycles) which account for $61 of the American consumer's expenditure, against $3 for that of the British; operation of transportation equipment—$63 in the U.S., as against $6 in the U.K.; purchase of clothing

and household textiles—$139 in the U.S., only $73 in the U.K.; and "miscellaneous" expenditures which consist partly of tourist expenditures abroad ($81 compared to $19). There are several other differences, but it is quite clear that about 60 per cent of the difference between the U.S. and the U.K. level of living is accounted for by the prevalence of consumer durables and semi-durables in the American consumer expenditure pattern. The differences become, of course, much more pronounced in comparison with the continental countries.

A counterpart of these differences in the expenditure pattern is to be found in the pattern of purchasing power equivalents. The purchasing power of European currencies is highest for commodities which the European consumer buys in larger, or in not much smaller, quantities than the American consumer; it is lowest for those items of which the former buys little and the latter much: cars, gasoline, household goods, clothing, etc.

These findings are by and large in accordance with a priori expectation. There is, however, one price differential which is more difficult to account for and which may be of great significance for the long-term growth prospects of the European countries compared with those of the U.S. "On the United States pattern of expenditures, purchasing power in European countries is about the same for both consumption and investiment, while on the European pattern of expenditures, investiment, except in Germany, is more expansive than consumption *In other words, to the extent that relative prices affect the allocation of resources, the European price structure is more favorable to consumption and less favorable to investment than is the case in the United States*" (p. 33—italics mine).

As the quotation and what has been said above about the "content" of the difference in the level of living between the U S. and the U.K show, the study contains a wealth of essentially new information which deserves to be analyzed to bring out its broader implications. Apart from a few isolated remarks, the authors have confined themselves—probably rightly so—to the presentation of statistical results It is to be hoped that they themselves and others will follow up the leads of the study.

About half the book is taken up with a detailed account of the methods by which the quantity and price data were gathered—the authors received what must have been most valuable help from the Organization for European Economic Cooperation and the Economic Cooperation Administration-Mutual Security Administration staffs—and by which the numerous problems of comparability, inadequacy of information, etc , were met. There are several paragraphs dealing lucidly with conceptual difficulties (e g , the pricing of government services, the handling of quality differences), and, on occasion, the authors take issue with other writers who have discussed problems of international comparability of national accounts, or, of national data generally. The account of the techniques used in preparing the data, and the comments on the broader issues of comparability, valuation, definition of end-products, etc., are refreshingly common-sense, and must impress anyone

who in his own work has faced similar problems, and has had to make similar value judgments. Welfare economists, both orthodox and sectarian, are bound to have misgivings about, and to take exception to, the free and easy way in which the authors have gone about tackling their problems and finding solutions where the welfare iconoclasts would deny the possibility of a solution. But those who prefer approximate common-sense answers to their common-sense questions will be grateful to the authors, and hope that the Organization for European Economic Cooperation will continue this work.

Individuals' Saving: Volume and Composition. *Irwin Friend with the assistance of Vito Natrella.* New York: John Wiley & Sons, Inc., 1954. Pp. x, 288. $5.00.

EDWARD F. DENISON, *Department of Commerce*

THIS is a most useful book which anyone concerned with interpretation of saving statistics will wish to read and to have available for frequent reference. Lucid and concise writing enables the authors to present in limited space nearly everything of importance that is known about individual saving in this country since the twenties and, of at least equal significance, to indicate what is not known, and why.

The range of subject matter is broad. It includes the meaning of personal saving and its conceptual relationship to business and government saving; description and appraisal of aggregate personal saving series derived by the three available methods—disposable income less consumption, investment less non-personal saving, and changes in individuals' assets and liabilities—and of saving and related statistics which have been obtained by surveys of households; and an adjustment of the results of several of the household surveys to place them on a conceptual basis comparable to the aggregate data. The household survey results are found to depart widely from the aggregate series, which are in fairly good agreement among themselves. Trends in both aggregate individual saving and its components (in terms of asset and liability changes) are examined, with some consideration of long-term trends but greater stress on more recent, and particularly postwar, experience. There is a careful analysis of both household survey results and time series data to ascertain what relationships can be shown to exist between saving rates and characteristics of the population and other economic variables; included is a critical analysis of deductions drawn by other investigators. Consideration of a variety of historical statistical relationships between saving and disposable income and other variables, their a priori reasonableness, and their effectiveness in forecasting postwar saving, concludes the main part of the study.

The last 123 pages of the volume consist of a greatly needed detailed description, not previously available, of the SEC data on individuals' saving, including a great deal of worksheet detail underlying the published estimates.

In the present state of knowledge, interpretation of saving statistics inevitably requires exercise of a great deal of judgment and understanding.

and even of intuition. The authors of this volume, from their long experience with the preparation and analysis of government data on saving, are unusually qualified to express worthwhile opinions on what reasonably can and cannot be deduced from the available information. While exercising proper restraint and, for the most part, providing the reader the necessary underlying evidence, they fortunately have not hesitated to express such views. With most of their judgments the reviewer would concur.

One which seems questionable is that the aggregate saving estimate derived by measurement of changes in individuals' assets and liabilities is probably more reliable than the other two aggregate series Most components of this estimate are obtained as a double residual Individuals' assets or liabilities at the beginning and end of a period are estimated by deducting estimates of the assets or liabilities of governments, corporations, and foreigners from those for the economy as a whole; the change in individuals' assets or liabilities during the period is then obtained as the difference between the amounts outstanding at the beginning and end of the period. Other components, for which the change is estimated directly, include items for which data are particularly tenuous It is by no means clear why this series should be preferred to the other two which, though also residuals, are derived from much smaller gross figures.

The volume is unusually free of obscure, slippery, or contradictory passages. One exception should be mentioned, however. On p. 99 it is said that on the basis of Kuznets' data "it cannot be stated with any confidence that there was or was not a moderate upward or downward trend [since the eighties] in the ratio of individuals' saving to income." On p. 100 it is concluded from "aggregate and survey results" that "there does not appear to have been any substantial upward or downward trend in the ratio of individuals' saving to income." Unless stress is placed on the difference between "moderate" and "substantial"—neither term is quantified—this is a very different statement The short intervening discussion of survey results, it may be noted, lends little support to the second conclusion, and particularly so in view of the authors' low opinion of the quality of the survey data.

That appraisal of consumer survey results, and an equally dismal view of the quality of nearly all pre-1929 saving statistics, may also call into question the authors' use of survey data to ascertain saving patterns and of individual saving aggregates for the twenties in the derivation of relationships of saving to income and other variables. They are utilized cautiously, however, and presumably only in ways the authors consider warranted.

The most general question the reviewer would raise about the discussion is whether the behavior of individual saving can satisfactorily be analyzed in isolation from that of business saving. As indicated in my article in the January 1955 *Survey of Current Business*, the record from 1929 to 1954, at least, suggests the presence of a more stable relationship between total gross private saving and total income or product than can be established separately between personal saving, corporate saving, and capital consump-

tion allowances, on the one hand, and relevant income series on the other. This at least raises a question as to whether changes in individual saving can be adequately analyzed without taking account of offsetting shifts in the other components of total saving. It may be noted also that it is the total saving relationship which is required for business cycle analysis, the parts being of less interest unless they help in appraising the behavior of total saving. It is no criticism of a study, however, to note that it is confined to its stated scope, and Friend and Natrella have rendered a valuable service in covering individual saving so adequately.

A Century and a Half of Federal Expenditures. *M. Slade Kendrick, assisted by Mark Wehle.* New York: National Bureau of Economic Research, Inc., 1955. Occasional Paper 48. Pp. x, 102. $1.25. Paper.

HENRY C. MURPHY, *International Monetary Fund*

THIS little booklet is an excellent reference work on Federal expenditures from 1794 to 1952. Kendrick and Wehle deserve the thanks of all persons interested in these statistics for their laborious and painstaking adjustment and classification of the underlying data, which are taken from a great variety of sources, and for the convenient form in which they have presented their results.

The booklet contains annual figures of total federal expenditures for the whole period expressed (1) in current dollars; (2) in dollars of constant (1926) purchasing power; (3) in 1926 dollars per capita; and (4) as a percentage of gross national product (before 1870 of national income at factor cost). In addition, the data are broken down into "military," "veterans," "interest," and "civil" (i.e., residual) expenditures for the whole period. Beginning in 1915, the additional category of "foreign" expenditures is added. The authors have done their best to make these classifications reasonable. In general, they have done a very good job. The only important exception which occurs to the reviewer is that, in his opinion, it would have been more realistic to have included the expenditures on the development of atomic energy under the head of "military" rather than, as they have in fact done, under the head of "civil."

The story of the great increase in federal expenditures over the period as a whole and of the tendency for these expenditures—both military and civil—to rise in steps following each great war (as compared with the prewar period) is, of course, a familiar one. It is natural, therefore, that the reviewer's attention should turn to special problems in the statistics. The most notable of these applies to the data on total Federal expenditures as a percentage of national income during the first half of the nineteenth century. These percentages are based on the underlying estimates in Martin's *National Income in the United States, 1799–1938* (National Industrial Conference Board, 1939) although, according to Simon Kuznets, cited by the

authors of this study, Martin's estimates have a downward growth bias and imply a *decline* in per capita real income between 1800 and 1840. It seems to the reviewer that such estimates provide shaky denominators for the ratios used in this study and that this matter should be thoroughly reinvestigated.

In discussing the growth in military expenditures, the authors present a wealth of data showing that both the proportion of the eligible male population in the armed forces and the expenditures per capita of the armed forces have increased substantially over the whole period—although the latter have increased much more than the former. In connection with the former, it seems to the reviewer that the age bracket 20–39 years, which is used as that most representative of the universe of all males eligible for the armed forces, is too high. It would scarcely stand the test of the Marine Corps obstacle course at Parris Island In connection with the latter, the authors have included an excellent appendix on "Technological Advances in the Weapons and Equipment of the Armed Forces." This covers within a brief compass the major developments in this field over the entire period and the reviewer found it fascinating.

Big Enterprise in a Competitive System. *A D. H Kaplan*, with a preface by Robert D. Calkins. Washington: The Brookings Institution, 1954. Pp. xii, 269. $4.00. Paper.

George H. Hildebrand, *University of California, Los Angeles*

Each generation seeks to overthrow the beliefs of its predecessor. In doing so it erects a new orthodoxy. Intellectual history thus exhibits an endless dialectic of interpretation and reinterpretation of human experience.

Currently, we are in the midst of a revisionist movement concerning the role of big business in the American system. Equally sincere observers in the 'thirties saw bigness as a major cause of inefficiency, rigidity, and secular stagnation. Today the votaries of revisionism in this field view the large firm as the mainspring of economic progress, fostering new dimensions to competition, accelerating the rate of innovation, and contributing to an unrivalled advance in the standard of living That this profession of faith is but a repetition of the forgotten gospel of the New Era stays these prophets not, though the recollection may be embarrassing to some of them

I do not think I do him an injustice when I include Kaplan in the revisionist school. As he says himself at the end of his book, "Big business has not merely been kept effectively subject to a competitive system; on the whole it has made an essential contribution to its scope, vitality, and effectiveness" (p. 248). His argument is reasonable, tempered by concessions to skeptics, and replete with many facts. How, then, did he arrive at this conclusion?

Kaplan poses his problem in two parts: (1) What is the relationship of big business to the changing structure of the economy; and, (2) How are we

to interpret its performance? He attacks the first question largely as a statistical problem, relying heavily upon already well-worked materials, augmented by a new survival test of the 100 largest industrials, 1909–1948. Concerning performance, Kaplan depends upon analytical assertions, illustrated here and there by fragmentary evidence. What emerges is an implicit and relatively untested, not to say optimistic, theory of big business, more likely to impress the already convinced than to convert the unconvinced.

Regarding big business and the economic structure, Kaplan looks first at the place of large firms in the total business population, using size distributions based upon number of employees. Here he finds that "the giant firm has grown with the economy but has had no measurable effect on the distribution of total economic activity as measured by employment" (p. 72). Next he turns to concentration of industrial production. Within the twelve largest manufacturing industries, plant size shows remarkable variance, with technology the principal influence. This holds true also for company concentrations by industry. Neither measure yields a definitive index of market power, Kaplan believes, because "industry" categories oversimplify, freeze, and distort actual competitive relations. Nor do measures of concentration by product turn out much better, partly for the same reason and partly because structural cross-sections conceal competition by innovation.

So Kaplan contends that concentration data must be supplemented by case studies—surely a proper inference. Less persuasive, perhaps, is his own foray into this field, particularly his *résumé* of the aluminum story, where he concludes that " . . . in the final analysis the price has been determined competitively in the market place" (p. 99). The only basis for this assertion is cross-product competition with other metals. If so, one may ask, as George Stocking does, whether market behavior was in any way modified by government-sponsored entry of two additional primary producers and by court decision in 1945?

Kaplan next considers the relative financial position of the large corporation—not because such measures are to him a reliable index of market power, but because financial scale commonly is so viewed and indeed often considered cumulative. Yet, whatever the measure, he finds no agglomerative trend favoring the large firm—by share of total corporate assets, corporate profits, or national income. Kaplan makes no attempt to examine the question of corporate interlocks beyond a later assertion that banker control has given way to professional management.

He breaks new ground with his test of turnover among the 100 largest industrials, 1909 to 1948. Of the original 100, 36 remained on the list at the end of the period. There were 47 newcomers in 1919, 28 in 1929, 11 in 1939, and 20 in 1948. So Kaplan concludes that turnover is high among the mighty—the top is a greasy pole, to borrow from Disraeli. Or in Kaplan's words, "There is no reason to believe that those now at the top can remain there any more than did their predecessors, short of alert participation in continuous product and market development" (p. 142). Perhaps so, but be-

fore we accept this conclusion it would be of interest also to know what role mergers played in the creation of these new leaders.

Regarding the place of large firms in the economic structure, Kaplan thus finds no causes for concern, either at the present time or on the basis of past history. He devotes the remainder of his book to an assessment of the performance of big business.

Here we find an ably written interpretation of the place of size in a competitive system, along the now familiar revisionist lines. Prices, for good reason, may be somewhat less flexible in the short run, but quality variation and innovation generally prevent realization of the gloomy theoretical vision of static oligopoly with shared markets. Where the struggle to attain and to hold positions is seemingly endless, "live and let live" is presumably a poor philosophy. In the large concern competition acquires a multiform aspect, reflecting internal pressures and rivalries, fickle and insatiable consumers, and a never-ending series of innovations with their enforced adaptations. Integration in all of its forms may lead to increased efficiency rather than to monopoly power, as at the turn of the century.

Kaplan concedes that difficult problems for public policy will still emerge, in certain types of discrimination, in "undue" leverage upon buyers and sellers. However, there is little that is new in his account. While calling for a "flexible policy" guided by a conception of reasonable limits in particular cases, he provides no concrete tests of what is unreasonable as to size, mergers, or market practices Integration as such, presents to him no serious general problems, and " has undoubtedly proved its capacity to promote expansion of the entire economy, including additional opportunities for smaller businesses and new users" (p 230). By contrast, and here surely he is correct, legislative attempts to enforce uniform prices have injured the common good.

The reassessment of the revisionists was no doubt overdue, but there is some danger that the reaction may go too far, though I do not intend this as a criticism of Kaplan in particular. Plainly the classical model of competition has serious shortcomings as a guide for interpreting behavior in a rapidly changing economy or for setting policy in that economy. Nonetheless, a satisfactory new theory has yet to be developed, if indeed one is possible.

As the foremost member of the new school, Schumpeter was probably right in suggesting that the rise of the large firm and the impressive advance in output were more than a chance coincidence. Yet before we proceed to relax comfortably on yet another "inevitable historical trend" we might well recall that large enterprise can also lead to stagnation, as much European experience indicates. In general the American story has been otherwise, given the differences in our cultural tradition and, in the environment in which that tradition has happily found expression. Yet we have some record of anti-competitive behavior here in the United States. Vigilance, rather than complacency, is the price of safety.

Capital and Output Trends in Manufacturing Industries, 1880–1948. Occasional Paper 41. *Daniel Creamer*. New York: National Bureau of Economic Research Inc., 1954. Pp. 104, $1.50.

KENNETH A. BOHR, *International Bank for Reconstruction & Development*

THIS paper is one of a series describing research being conducted by the National Bureau on long-term trends in capital formation in the major capital-using sectors of the U.S. economy. It presents concisely yet with ample documentation the major findings for manufacturing industries. In a long introduction Simon Kuznets places the paper in the perspective of the larger program of research and speculates briefly on some possible causes for the observed trends.

The analysis is centered around the behavior of the ratio of capital stock to output (the capital-output ratio) for a series of benchmark years (1880, 1890, 1909, 1919, 1929, 1937 and 1948) selected on the basis of availability of data and similar relative positions in the economic fluctuations of the period. Data on capital stock were obtained from Censuses of Manufactures for the period 1880–1919 and from corporate income tax returns for 1929 and thereafter. Measures of output were obtained from the same sources. The principal finding is that a significant rise in the capital-output ratio for manufacturing industry as a whole took place from 1880 to the decade 1909–1919 followed by a definite and substantial decline to 1948, the most recent date studied. In other words although capital stock increased throughout the period its rate of growth exceeded that of output up to around 1919 and was exceeded by output's rate since then, at least until 1948. This observed trend was tested in various ways. It was found that it was not affected by varying the concepts of capital and output used. It held when either total capital, fixed capital or working capital were used and it held when output was taken gross or net. It was also found that neither the changing structure of industry nor the changing size of firms offered adequate explanations of the trend.

In addition to the main findings on manufacturing industry as a whole mention should be made of the interesting material presented on the behavior of various groups of industries—material collected in order to investigate the change in industrial structure over the period and the effect of this change on the over-all ratio. It was found possible to distinguish 41 minor industries classified into 15 major groups over the entire period. The distribution of total capital stock was computed in terms of these 41 industries for each benchmark year and capital-output ratios were determined for each industry for each year. The detailed picture this data presents of the changing structure of industrial capital, changing capital-output ratios over time, and of the marked differences in the levels of the ratios among industries represents a considerable amount of new knowledge that should certainly prove most useful to students of industrial growth.

No attempt is made to explain the findings. However, various suggestions

of possible explanations are made by Creamer, and by Kuznets in his introduction. Both mention that the spread of technological innovation of a "capital-using" type may help explain the early rise in the capital-output ratio and Creamer also suggests that extensive industrial development as shown in the rapid growth in the number of establishments during this period may have also been a factor. To account for the later decline in the ratio technological change is also mentioned as well as the usually assumed drive of the entrepreneur to use scarce resources more efficiently. And on a more specific level Kuznets suggests that with a decline in the rate of growth of gross fixed capital formation the per cent of current capital stock renewed tends to grow and the renewed stock, generally being more productive than that which it replaces though costing no more in constant price terms, tends to depress the ratio.

It is of course a long way from a suggested hypothesis to a tested explanation. However, a forthcoming monograph which promises to deal with the same subject matter in more detail may be expected to make the task of explanation somewhat easier. So this excellent study leaves one stimulated, informed and eagerly awaiting the promised sequel One might add that in these times, when economists are increasingly preoccupied with problems of economic growth and when the concept of the capital-output ratio is paying for its current popularity with a certain amount of indiscriminate use, it is particularly satisfying to see a competent study such as this supply a much needed perspective to the concept, at least insofar as it refers to one important economic sector.

Canada and the International Business Cycle, 1927–39. *Edward Marcus.* New York: Bookman Associates, 1954 Pp. 211. $3.75.

R. Craig McIvor, *McMaster University*

This case study was designed to narrow a serious theoretical and empirical gap in the literature of the business cycle, i.e., the problem of how cyclical influences are transmitted internationally. Specifically, it is concerned with the interrelationships among Canada's domestic activity, foreign trade, and international capital movements, as they have reacted and adjusted to forces originating abroad—whether in response to the business cycle of an important customer (e.g., the United States or the United Kingdom), or the workings of international competitive forces.

The first two chapters provide a summary economic background and an analytical framework from which the detailed study of cyclical forces is subsequently developed. Thus, such features of the 1927–29 Canadian economy as its primarily extractive basis, its relatively great dependence on foreign trade and the unique geographic distribution of that trade, vis-à-vis the United States and the United Kingdom, are noted. Also mentioned are the lack of a central bank and of discretionary monetary control before 1935, the *de facto* departure from gold at the end of 1928 and subsequent

exchange fluctuations, and finally, the shift in Canada's international position from a net capital importer to exporter during this period.

It is asserted that the Canadian business cycle is subject to two groups of external influences, i.e., *forces of a general economic* nature and those of a *cyclical* nature. The former, which include such variables as changing tastes, customs, trade barriers and competitive pressures, are not produced by any cyclical movement abroad, while the latter are directly related to, and would not occur in the absence of, such external cyclical movements. It is the author's basic contention that the external determinants of Canada's downturns in 1929 and 1937 are to be found in the *general* economic forces just noted, but that once her economy became depressed, Canada was powerless to initiate a recovery, this depending upon a prior upturn in the United States and the United Kingdom, i.e., upon external *cyclical* influences. Summarizing this thesis, Marcus claims that "somewhat facetiously, one might say that Canada enjoyed enough independence to get herself into trouble, but that she needed outside help to emerge from the resulting depression."

The detailed analysis of the Canadian fluctuations is divided into four periods: 1927–29 (boom and collapse), 1930–33 (recession and recovery), 1933–37 (recovery) and 1937–38 (recession and recovery). The end of the Canadian boom during the late 'twenties is found to have been foreshadowed by sagging exports of woodpulp, newsprint, wheat, and flour in 1927 and 1928. The adverse effects upon income "rippled" through the economy with cumulative force, and the downturn occurred several months before the comparable development in the United States. These depressing influences are described as "exogeneous forces of a somewhat random nature" independent of any cyclical influences abroad.

As long as Canada's two main export markets, the United Kingdom (wheat and flour) and the United States (woodpulp and newsprint) remained depressed, recovery was impossible, but following their respective upturns in 1932 and 1933, the Canadian economy emerged from its period of stagnation to one of steady expansion. This reversal is found to have been clearly dependent upon cyclical movements abroad. In the case of the 1937 recession, Canada's wheat and paper export indices faltered several months before the actual business downturn, "in a pattern strikingly similar to that of 1927–29." Once again the explanation of this faltering runs not in terms of cyclical movements in the United States and the United Kingdom, but rather of a "fortuitous" combination of random occurrences. When World War II arrived, the Canadian economy had once again entered a period of expansion, for which the main stimulus was provided by rising exports related to the European armament program, clearly a non-cyclical influence.

The responses of the Canadian balance of payments to external disturbances during the whole period are alleged to have differed from what might reasonably have been "expected" (gold standard conditions being assumed).

Marcus finds that during the economic contraction following 1929, imports actually fell more than exports, an unfavorable balance of payments became favorable, and net export of capital developed in the early 'thirties. Conversely, the boom of 1937 was accompanied by a rapid deterioration in Canada's current account, and it is generally regarded as surprising that "the worse the state of domestic activity, the better the international position became." Two brief comments may be made on this generalization The first is that it was simply not the case during the upturn after 1933; the second is that the author, instead of rejecting the foreign trade multiplier as an irrelevant concept, might have incorporated it into his analysis, recognizing how its operation might be modified by "particular" circumstances.

One of the most interesting sections of the study deals with the well-known limitations of fiscal policy in an open economy. It is not surprising that Canadian government deficits are found to have exerted little influence upon the course of economic affairs during the 'thirties For the most part, these were "reluctant" deficits, for the philosophy of counter-cyclical financing can scarcely be said to have gained serious acceptance in Canada before World War II.

In any such empirical investigation, it is easy to raise questions both of fact and of analysis and interpretation. In the former category, we have already noted Marcus' implication of the negative correlation between national income and the current account balance after 1933; since the wheat "holdback" plays an important part in the analysis of the 1929 decline, it is not irrelevant to point out that it appears to have occurred about a year later than suggested. Concerning analysis and interpretation, the out-of-hand rejection of the foreign trade multiplier may be seriously questioned. The reviewer would also have welcomed a more explicit assessment of the relative importance of domestic versus external factors in Canadian cyclical fluctuations. At various points, there are unqualified statements whose validity is surely open to serious question. Thus (p. 179), with reference to the early 'thirties, "the use of exchange depreciation was also circumscribed, thus preventing the easiest means of stimulating exports." A plausible argument of many years' standing suggests that depreciation, in the particular Canadian circumstances of the time, might have had precisely the opposite effect. Again (p. 188), with reference to the interest rate as an instrument of economic control, "Canadian economists feel that the influence of the rate of interest is not too great." Granted that such a statement has little meaning if taken literally, one may still question its general validity, in the light of recent monetary experience.

Such criticisms should not obscure the fact that a great deal of valuable empirical research is contained in this detailed study of the Canadian economy, and no student of cyclical fluctuations can fail to benefit from a careful reading of this monograph.

The Economics of Recession and Revival: An Interpretation of 1937–38. *Kenneth Roose.* Yale University Studies in Economics, Vol. 2. New Haven: Yale University Press, 1954. Pp. 280. $4.00.

BERT G. HICKMAN, *Council of Economic Advisers*

This is an ambitious book. Roose has set for himself the difficult task of assessing the relative importance of various causal factors in the contraction of 1937–38. Among other factors he discusses monetary and fiscal policies, price-cost relationships, changes in the degree of monopoly power, underconsumption, shortage of capital, and the political and social turmoil of the period. The opinions of other writers, methodological issues, and the implications of the study for the formation of public policy also receive attention. Roose is primarily concerned with the causes of the downturn of 1937, but the upturn of 1938 is treated briefly and at various points he addresses himself to the question of why the entire expansion from 1933 to 1937 was so lacking in vigor. Unfortunately, his ambitions are not fully realized and his major conclusions do not carry complete conviction, partly because of the complexity of his subject matter and partly for other reasons.

One source of difficulty may be mentioned in passing. The presentation of statistical data is such that it is sometimes quite difficult for the reader to follow the analysis and to check the judgments of the author—an important drawback in a study of the quantitative-historical type. Most troublesome in this connection is the frequent introduction of data into the text in the form of values on two or more comparison dates or of percentage changes between such dates. Where data are not shown as continuous time series it is often impossible for the reader to answer questions that come to mind concerning variations in rates of change during the intervals between comparison dates, time-lags, and the like, and the force of the author's argument is inevitably reduced.

The study is organized around a national income framework in the sense that attention is focussed on the major income and expenditure flows and developments in the markets for money, wages, and goods are viewed from the perspective of their possible influence on investment and consumption. This does not mean that use is made of formal aggregative *cycle* models; on the contrary, Roose is often interested in hypotheses and variables which cannot be fitted easily into such models, and his chapter on methodology stresses the uniqueness and complexity of each historical business *fluctuation*. What it does mean is that wherever possible a given causal hypothesis is tested against both the observed movements of the variables directly involved and the relevance of those variables to income and expenditure flows. In this connection, it is worth noting that each hypothesis is tested in turn. An advantage of this procedure is that it facilitates the formal testing of alternative theories, but it also carries an important disadvantage. It tends to direct attention away from possible inter-relationships between developments in various sectors of the economy.

Let me illustrate. In the chapter on fiscal policy, the author argues that

the sharp decline of "net government contribution to income" in the first half of 1937 was not the primary cause of the downturn on the grounds that personal income did not decline until the third quarter and that income-expenditure lags could not account for such a long lag. Although he points out that the bulk of government expenditures of that period were quickly received as income, he fails to note at this point that Barger's series on adjusted income and consumer expenditures decline sharply in the first quarter and rise only moderately in the second quarter. Later, in the chapter on consumption, "underconsumption" is rejected as an explanation of the downturn, primarily on the grounds that "Only the quarterly marginal propensities to consume show consumption lagging, and even here it accelerated immediately prior to the recession" and that "consumers' income was apparently not outdistanced by the production of consumers' goods . . ." However, Roose's calculations show that real adjusted income, real consumption, and consumer goods production all declined between the last quarter of 1936 and the second quarter of 1937, and that consumer goods production continued to fall even when income and consumption accelerated briefly in the third quarter. Still later, in the chapter on inventories, the judgment is reached that although inventory investment increased quite substantially during the first three quarters of 1937, inventories did not become burdensome until income and expenditures declined, and were therefore a secondary factor in the recession. This judgment is reached despite the fact that he earlier concludes that inventories were excessive relative to new orders in the summer of 1937 and the further fact that new orders declined substantially throughout the first half of the year.

It is not implausible that the various developments just described were causally related and were of considerable importance in precipitating the downturn. That is, the decline in net government contribution to income may have induced the decline in consumer expenditures and the latter in turn may have been responsible for the decline in consumer goods production and in new orders during the first six months of the year. Under the circumstances, it is difficult to accept without reservation Roose's judgments that the slackening of consumer expenditures in the first half of the year had little to do with the recession or that rapidly mounting inventories were not considered troublesome enough to lead to efforts to reduce them until late in the year. Perhaps the hypothesis I suggest would not stand up under careful empirical investigation, but the point that I wish to stress is that it is not investigated at all, quite possibly because each theory of the turning point is discussed independently. A procedure wherein alternative hypotheses are tested independently and those which are judged consistent with observed behavior are retained and the others rejected may not yield the same set of causal factors or the same judgments of relative importance as a procedure in which hypotheses are considered as possible complements as well as competitors.

Roose concludes that three major elements were responsible for the down-

urn. The decline in net government contribution to income shifted the responsibility for sustaining and increasing aggregate activity to the private sector. The new undistributed profits tax reduced the liquidity of corporations and the restrictive monetary policy of the Federal Reserve raised the cost of capital, while both actions adversely affected business expectations. 'Most important of all, however, was the reduced profitability of investment, beginning in the first quarter of 1937. This resulted from the increased costs, in which labor costs played a prominent part." In my opinion, a Scotch verdict of "not proven" should be returned on this last proposition—the data with which to establish either guilt or innocence are simply not available. As Roose himself points out, the existence of a price-labor cost squeeze annot be clearly demonstrated by comparisons of changes in average hourly arnings and wholesale prices. Data on labor costs per unit of output, which ake account of changes in productivity as well as in wage rates, would seem be a minimal requirement for this purpose. Moreover, average hourly arnings did not increase sharply relative to prices of finished goods until he second quarter of 1937, or the very quarter in which prices of raw materials and semi-manufactured goods began to decline, so that higher wage ates may have been offset at least in part by falling prices of materials.

---

merican Agriculture: Its Structure and Place in the Economy, A Volume in the ensus Monograph Series. *Ronald L. Mighell.* New York: John Wiley & Sons, 955. Pp. xii, 187. $5.00.

VARDEN FULLER, *University of California, Berkeley*

THE Director of the Bureau of the Census and the President of the Social Science Research Council in the foreword to this volume explain that it part of an effort to regenerate the analytical census monographs which, ecause of depression and war, have declined since the Census of 1920 he publisher says that it is one of a series "of interest to economists, agricultural experts, statisticians, political scientists, and other social scientists." contrast, the author's preface states that the book "is addressed to readers o are interested in agriculture, but who are not necessarily professional ricultural specialists. Its purpose is to present in ordinary words a picture the structure of agriculture. . . . " Thus, it would appear that the book as launched and produced under divergent conceptions of purpose

The major misfortune of the book seems to lie in the effort concurrently serve these essentially different purposes. Moreover, this effort is made e more ineffective by the discussion of questions not likely ever to be asked the nonspecialist and in reference to which the specialist would immediately seek more basic sources and exhaustive treatments. Statistical d conceptual problems in measuring scale of farm enterprises and their spective efficiencies, and in defining and measuring types of farming and plaining their locations are prominent examples.

The work offers very little material based on special tabulations or in-

tensive analyses of census data. Indeed, the Census has no outstandingly prominent role; regularly published works, agricultural college and Department of Agriculture bulletins, and estimates made by the author's colleague-specialists in the Department of Agriculture share the burdens of source and authority. Most of the chapters deal with orthodox subjects—scale of farms, types of farms, farm tenure and debt, social features—that have traditionally been found in texts on farm management, land economics, rural sociology, or agricultural economics. In respect to these subjects, there is some recent information, particularly on trends and tendencies of change, but less new analysis and virtually nothing new for the specialist who has kept informed.

The chapter on "Agriculture in the Total Economic Process," a significant subject of growing interest and on which the 1950 Census obtained relevant information, occupies 6½ pages that innocently discuss the percentage of population on farms, the percentage of national income that originates in agriculture, the farmer's share of the consumer's dollar, the expenditures of farmers for production goods and concludes that a farm is a biological manufacturing unit that is interdependent with other industries. Curiously, input-output relations are not treated under this heading but in a later chapter called Group Interests in Agriculture, and there only to the extent of a short abstract from Leontief's *Scientific American* article. In this context, the author employs input-output not to illuminate intricate economic relations but to help convince one (actually, Adam Smith no less) that the twentieth century is very different from the eighteenth and that in consequence "he would begin to understand why a labor movement and a farmers' movement have arisen to provide countervailing power for these groups."

Again, in a chapter on "Farm Tenure and Debt," the author picks up, then drops in two sentences, the subject of contractual production. Although this is a direction of rapid and significant expansion with far-reaching implications for production, processing, and marketing, we are inconclusively advised in the second of these two sentences that: "Contracts take many forms, the processing company goes so far in some instances as to prepare the land, supply the seed, and harvest the crop."

Another aspect of "Agriculture in the Total Economic Process" that concerns both the specialist and the nonspecialist is incomes and income source relations. Aspects of this subject are picked up in several chapters but nowhere definitely or unambiguously discussed. In Chapter 2, percentages of population on farms are set up against percentages of national income from agriculture but with no explanation that farm population and the population engaged in agriculture are not identities. In Chapter 3, we are told, strangely, that "hired labor is paid for as in other occupations." In Chapter 4, the text says (p. 55) that earning money from agriculture is the primary interest of all commercial farm operators whereas the footnote says it is not; more lamentable is the fact that the 1950 Census endeavored to collect information that would throw some light on the complexity of occupational

relations between agriculture and nonagriculture, of which the author makes virtually no use. Chapter 9 again picks up the subject of incomes and their magnitudes by classes of farms, and although nonfarm income received by farm families is included here, its interoccupational significance is lost in the emphasis on income differences within agriculture.

The book unfortunately is well supplied with evident contradictions and non sequiturs. In Chapter 1, some soliloquial-like sentences wondering whether what has happened in agriculture in the past half century should be called "change," "trend," or "development" lead abruptly to invoking the authority of T. W. Schultz to pronounce there is no received theory explaining economic development Similarly, a series of sentences (pp. 24–25) which declare that modern day technology decreases the rigidity of our bond with the soil terminate with: "In the Western World the fear of Malthusianism recedes into the background." Those who pay attention to what they have read from one paragraph to another will have to resolve for themselves on such matters as whether hybrid seed corn is expensive or inexpensive (pp. 14, 24), whether man has attained control of his physical environment or not (p. 41), whether or not the "agricultural ladder" really exists and if so can it be climbed (p. 95), whether tenancy is good or bad (pp. 98, 101).

A book with so pretentious a title and so well sponsored ought either to be a useful monographic source for the specialist or an enlightening and unambiguous descriptive survey for the general reader and the undergraduate student. This is neither.

Factor Analysis. An Introduction and Manual for the Psychologist and Social Scientist. *Raymond B. Cattell.* New York: Harper and Brothers, 1952. Pp. xiii, 462. $6.00.

HERBERT SOLOMON, *Teachers College, Columbia University*

IN THE preface Cattell states that this book has been written to meet three principal requirements. "First it sets out to meet the need of the general student in science to gain some idea of what factor analysis is about and to understand how it integrates with scientific methods and concepts generally. Second, it is intended as a textbook for statistics courses which deal with factor analysis for the first time, either as an appreciable part or as the whole of the semester course. The third objective is to supply a handbook for the research worker, the student, and the statistical clerk which will be a practical guide with respect to carrying out the processes most frequently in use." The book is divided into three sections, each designed to treat one of the three aims. On the whole, Cattell does very well by his goals and makes a valuable contribution to the literature of factor analysis.

The book is written wholly in the framework of the multiple factor structure; a model which states that a response is a linear function of factors. Where the book treats this model as a research technique applicable to many different disciplines, much more general use can be made of it than where it

treats the model as actually representing a behavioral pattern leading to observed psychological responses. The statistician can read this book without too much concern for the psychological implications of the model He can borrow many of its procedures for use in other disciplines, but the statistician should be aware that, as the reviewer wholeheartedly feels and the author states in his preface, "Factor analysis requires the skills of an art as well as statistical principles "

Thurstone's centroid method of estimating parameters in the multiple factor structure is featured prominently in the text. Very little space is devoted to Hotelling's principal components method of estimating the parameters, and this is to be expected if the "simple structure" concept is to be applied after factorization However, some psychology groups are still interested in principal component estimates, and it may prove useful in other disciplines There may even be a use for principal component analysis under the assumption of the "simple structure" criterion if the criterion is applied before factorization rather than in concert with the factorization.

Within the framework the author has set for himself, he writes lucidly, interestingly, and intelligently. Cattell gives a nice development of "alphabet soup" factor analysis, that is, the labelling of a factorization by the manner in which the elements of the initial correlation matrix are computed. For example · in R-technique, traits are correlated, in Q-technique, persons are correlated, in P-technique, traits on one person are correlated; in O-technique, correlations refer to correlations of occasions for a single person.

This book will not be as appropriate for the uninitiated student as the author thinks it is, nor will it be the manual for the statistical clerk that the author has strived to make it; but it is a valuable addition to the list of textbooks for courses in factor analysis and to the list of reference books for serious researchers

---

The Population of Yugoslavia. *Paul F Myers and Arthur A Campbell.* International Population Statistics Reports, Series P 90 no 5 U S Bureau of the Census. U S. Government Printing Office, 1954 Pp vi, 161 $1 00

Miloš Macura, *Yugoslav Federal Statistical Office*

This book appertains to the series of reports devoted to the population of foreign countries published by the Bureau of the Census Apart from 28 tables and 16 charts, the report contains 112 pages of text and 6 appendixes.

To this reviewer the tables are of the greatest interest. Most of them (24) are well selected from the prewar and postwar Yugoslav publications bearing on population census figures, vital statistics, school statistics and labor statistics. International statistics are represented by two tables, whereas the tables "Abridged life tables for males and females 1950" and "Estimates and projections of the population of Yugoslavia 1953–1970" have been prepared by the authors and their collaborators It should be stressed that the conversion of data to actual boundaries and states of Yugoslavia exacted a

comprehensive and meticulous work, successfully published now.[1] In doing so, the authors encountered certain discrepancies in methods and procedures successfully overcome in general.[2] In working out the population projection 1953–1970, the authors had to prepare the abridged life tables. It was not possible to ascertain whether they made use of the Yugoslav life tables[3] for the prewar period. The future number of the Yugoslav population computed on the basis of four assumptions (amounting from 20,983,000 to 22,987,000 of inhabitants for the year 1970), is discussed in the last chapter of the text "Prospects for population growth." Appendix F gives pertinent technical details. The reader gets the impression that this chapter is a most successful one as regards the methods and carefulness applied.

The first eight chapters of the text deal with: "Summary," "Introduction," "Population development," "Distribution of the population," "Demographic and social characteristics of the population," "Education," "Economically active population," and "Differential fertility." Apart from the interpretation of statistical facts, these chapters give surveys of historical, social, economic, cultural, political, etc., factors and institutions.

It is very difficult for an author without deep knowledge of conditions, institutions and language of a country to enter into such considerations with the purpose of giving meritorious opinions; and the authors of *The Population of Yugoslavia* faced just such a difficult assignment. It should be stressed and appreciated that they consulted some eighty sources, but less than twenty were Yugoslav ones. Nevertheless, the sources were insufficient for such a study, particularly owing to a very small number of Yugoslav original studies published in English.[4] The authors were not sufficiently aware of the influence exerted by the historical factor upon the shaping of the population of different regions of Yugoslavia. A very slow economic development and an agrarian overpopulation have been appreciated in rather cursory a manner and without necessary consideration of all the conditions of development. The postwar efforts made in the country's industrialization and economic progress are valued positively. However, this reviewer feels that the authors did not comprehend the whole complex of the economic problems of countries not advanced industrially.

The Yugoslav reader will ask himself how was it possible that the authors so far underestimated war losses of the population and overestimated the emigration; he will be particularly surprised at the overestimation of Ger-

[1] Two appendices are devoted to it: A. Former Italian territory ceded to Yugoslavia in 1947 and B. Relationship between the boundaries of the historical provinces, banovinas, and the present republics in Yugoslavia.

[2] See appendix C. Population Censuses of Yugoslavia, D. Differences between census and other population data, and E. Errors in reporting age.

[3] Lah-Ivo: Tablice za finančno in aktuarsko matematiko (Tables for financial and actuarial mathematics), Ljubljana, Državna založba Slovenije, 1947.

[4] In my opinion, some of smaller errors are not serious, as for instance, the first natality statistics were available from 1890 (p. 2), that Istra forms a part of Slovenia (p. 10), that Germans attacked Yugoslavia on April 1st, 1941 (p. 11), that natality and mortality rates are established in 1891 only (p. 17), in appraising internal migration (p. 30), in appraising a high sex ratio (p. 42), etc.

man minority emigration and at the underestimation of the mortality of the same minority taking part in the Nazi armed forces occupying Yugoslavia. The Yugoslav reader wonders also at the doubts expressed by the authors in evaluating the accuracy of the answers regarding religious conviction in the 1953 Population Census; and this the more so because the needed information was, or could have been made, available to the authors.[5] Remarks of similar kind may be made also with regard to the passages setting out in an inadequate manner the course of the recent history of Yugoslavia, etc.

Nevertheless, these critical remarks cannot diminish the value of the publication. As the first publication on the Yugoslav population appearing in English, it deserves full attention.[6] It contains needed information on essential problems of the development and condition of the Yugoslav population. As Yugoslavia is now in the process of rapid industrialization and general development, such a book is of interest not only as containing bare information, but as a basic text-book for forthcoming researches as well In this sense, tribute is to be paid to the Editor's initiative and to conscious and assiduous efforts of the authors.

Food Consumption of Urban Families in the United States, with an appraisal of methods of analysis. Agriculture Information Bulletin No 132. *Faith Clark, Janet Murray, Gertrude S. Weiss, Evelyn Grossman,* U S. Department of Agriculture, *Agricultural Research Service, Home Economics Research Branch*

Doris P. Rothwell, *U.S Bureau of Labor Statistics*

This 200 page bulletin is a useful addition to the library of reference materials on family consumption It presents results and also describes methods for their analysis and is well organized for the user of the data. Over half the volume consists of standard statistical tables presenting basic data on urban family food consumption for groups and items, during the period 1948–49. These tables alone would make this a valuable publication. It brings together in one volume results of the nation-wide survey of urban housekeeping families in the spring of 1948, surveys of housekeeping families in 4 cities (Birmingham, Ala., Buffalo, N.Y., Minneapolis-St Paul, Minn ; and San Francisco, Calif) in the winter of 1948, and seasonal surveys in

[5] It is characteristic, that the quality control of last census data omitted the verifying of the answers to the question on religion It was, namely, considered that such a question bears an eminently subjective character, so that it might not be submitted to any control Žarković, S S , Population Census errors, Beograd 1954, P Depoid, Rapport sur le degré de précision des statistiques démographiques, International Statistical Conferences, Rio de Janeiro, 1955

[6] It should be stressed that the Federal Statistical Office did not publish such a synthetic publication concerning Yugoslav population Its ten volumes *of the 1948 Population* Census results contain ten analyses of particular population characteristics. In doing so, the Federal Statistical Office bore in mind the *fact* that the 1948 Population Census, giving first postwar results, was a short census only By its nature, such a census was not in a position to furnish all the information needed to make a detailed population picture of Yugoslavia Such a survey is to be expected after the publication of all the 1953 Population Census results.

these 4 cities for additional periods during 1948–49. There is also a comparison of results with a similar survey in 1942.

In the text material the authors have made an effective separation of the technical details from the analytical summary presented in Part I. The summary is not extensive. However it contains a good discussion of important family characteristics associated with variation in consumption. Its analysis of changes in family food consumption from 1942 to 1948 also is interesting.

Part II, "Some Problems and Methods of Analyzing Family Food Data," contains an excellent discussion of the limitations of the data and the pitfalls for the uninitiated analyst. It points up the interrelationships of factors affecting consumption and describes various techniques for their segregation, including some illustrative examples, particularly for milk. It contains a good discussion of alternative methods of adjusting for differences in family size. Particularly noteworthy is the insertion of nontechnical descriptions for the technical terms used. For example, referring to income elasticity, the report states "income elasticity—a term for the ratio of rates of change in the consumption of an item and in income—is an indication of the order of urgency or degree of preference in consumption," and again "income elasticity may be defined as the relative change in quantity consumed (or in expenditures) divided by the relative change in income, other things being equal." The way in which consumption data by income group is adjusted for family size is very simply explained (p. 40). Only in the discussion of the more complex regression and correlation technique for determining factors affecting consumption (pp. 29–31) were the writers apparently unable to provide nontechnical explanations.

Part II also contains useful analytical statistics, such as tables of income elasticities and coefficients of variation for selected foods.

At the end of Part II is included a brief statement on seasonal indexes of food consumption. Because of the dearth of information in this field, the data shown are valuable as general indicators. However they must be used with caution since they were based on limited data and required important adjustments and assumptions. The report states clearly that they represent a single year, 1948, but the term "seasonal indexes" appears to imply a greater degree of stability from year to year than was intended.

Appendix B of the Bulletin, "Methods Used in Collecting the Data," contains an honest and complete accounting of the survey methods. Included also is a fairly comprehensive appraisal of the validity of the data. Unfortunately the fine print makes difficult reading of this valuable portion of the report.

The sampling methods used deserve some comment. The city sampling design provided for stratification of all urban places by geographic region and size of city. An extra restriction (within a small tolerance) of State distribution according to 1947 population was imposed, and random samples were redrawn for each stratum until the proper distribution was obtained

The report states that 4 to 50 drawings were required. In the block sampling for the 32 largest cities, restrictions on the average number of dwelling units and average rental value based on the 1940 Census were imposed, which required one to six drawings Statistically, another method (possibly better) would have been to build such restrictions into the sampling design through further stratification

This sampling procedure appears to be akin to "balancing" discussed in some detail by Yates in his book on sampling methods,[1] in which the sample selection is such that the average value of the sample units for some characteristic is equal to the known average for the universe In other respects the sample is equivalent to a random sample. Under Yates' method, balancing would be effected through replacement of individual units by random selection rather than by complete redrawing of a cell, as was done by the Department of Agriculture.

According to Yates, balancing by stratum should only be used when a "moderately large" number of units is selected from each stratum. Otherwise he says, "undue restrictions will be placed on the sample which will result in the selection of a sample which is not otherwise equivalent to a random sample " The point is important since in some of their analysis the authors of this bulletin have treated the sample as if it were random. As an alternative to stratification, the procedure would be effective only if the values under investigation (in this case, family expenditures for food) are roughly proportional to the differences in the known characteristic Otherwise Yates says, stratification would be preferable

The appraisal section of the bulletin considers the representativeness of the sample by comparison with Census data. As would be expected in view of the basis for stratification, data on characteristics of households in the sample agreed closely with Census data The report then discusses the sampling errors of the data on the assumption of a random sampling procedure. It contains an enlightening table of sampling reliability for various foods by income group. Next, the data were reviewed for consistency and for conformance to established consumption patterns. Finally a comparison was made with national estimates of the Department of Agriculture on per capita consumption. On all these counts, the sample appeared to be fairly reliable.

Throughout Part II and Appendix B the writers are careful to point out limitations of the data, but in addition they provide their own overall evaluation of their importance. For sophisticated users, parts of this will seem obvious, but for users not accustomed to analysis of consumer expenditure data, this section will be valuable

---

[1] Yates, Frank *Sampling Methods for Censuses and Surveys*, 1949 Also "Review on Recent Statistical Developments in Sampling and Sampling Surveys," *Journal of the Royal Statistical Society*, Vol 109, 1946

Life Insurance Ownership Among American Families, 1954, ***Survey Research Center.*** **Ann Arbor: Institute for Social Research, 1955. Pp. iv, 64. Free.**

ROBERT FERBER, *University of Illinois*

This monograph is another in the series of annual reports begun in 1948 which the Survey Research Center has prepared on life insurance ownership based on data compiled in its annual *Survey of Consumer Finances.* One question was asked on life insurance in the 1954 Survey, namely: "Do you carry life insurance?" followed by subqueries on premiums paid in 1953, and any coverage at all during 1953 if no premiums paid. The organization of the report is built around the answers to this question, one chapter comparing trends in life insurance ownership from 1949 to 1953, two others the relation of demographic and economic factors to insurance ownership, and a final chapter to a cursory exploration by means of cross-section tabulations of the interrelations between insurance ownership and these other variables.

Like its predecessors in this series, much useful and interesting material on life insurance ownership is provided by this monograph. Many people will undoubtedly deplore the absence of information on the types of insurance policies held, amount of coverage, and other characteristics, but considering the length of the original schedule and the subsidiary status of the general subject in the schedule, one cannot reasonably ask for additional questions on life insurance ownership. Within the framework of the present report, however, a number of questions can be raised regarding the organization and analysis of the material. Briefly enumerated, they are:

1. The basic analytical approach was to consider first and at some length the bivariate relations between life insurance ownership and various demographic and economic variables and then in a final brief chapter the effect on these relations of the interdependence between variables. An attempt is made to eliminate the interacting influence of income, but the result is nevertheless that the reader is told only at the end that certain relationships demonstrated to exist by means of two-variable cross-tabulations are in fact due to interacting effects of other variables. This is what may be termed the "orthodox" approach in such cases—to consider two-variable correlations, then three variables, then those of higher order—but its desirability seems to this reviewer questionable in a study such as this one. Not only does it tend to confuse statistical laymen, who probably constitute the bulk of the study's audience, but much of the preliminary material turns out to be at best unnecessary and at worst misleading. More desirable would be an approach that begins by explaining simply and briefly the problem of interactions, proceeds to examine the interactions in some detail and to derive the "true" relationships, and finally to consider the significance or implications of the findings.

2. Some attention might have been given to the type of audience for whom this monograph was intended. If intended mainly for statistical laymen, the presentation could have been improved considerably by using charts and other pictorial representation (which is completely absent) and relegating

many of the text tables to an appendix. If intended for statisticians, more powerful analytical methods might have been used, if only for supplementary purposes, instead of placing complete reliance upon examination of cross-tabulations. And if intended for both groups, a merger of the two preceding suggestions would not be at all difficult.

3. Sadly lacking is a summary of the main results, preferably at the beginning of the monograph. There is, however, a two-sentence summary at the end of the last chapter which neatly presents the gist of the study:

> The analysis presented in this report has shown that the two factors which are most important in determining life insurance coverage and premium payments are income and age. Other factors which are associated with insurance ownership and premiums are type of community, occupation, veteran status, liquid asset holdings, and payments to retirement and pension funds.

Unfortunately, however, this summary is so submerged in the text that only the rare person who reads the study from cover to cover is likely to stumble over it.

4. Hypotheses are conjectured here and there to account for certain observed relationships and left as such when a few extra cross-tabulations might have provided conclusive answers. Thus, on p 49 it is conjectured that most spending units making no social security payments and not insured are headed by retired persons Since occupation-of-head data were obtained, one further cross-tabulation would have tested this conjecture.

5. Might the question on which this report is based be conducive to underestimation of life insurance ownership? Spending units were classified as "insured" if premiums were paid on at least one member Unless special care is taken in the interview, however, the "you" in the basic question "Do you carry life insurance?" might easily be interpreted by the respondent as referring to himself alone.

6. Some more attention to organizational details would not have been inappropriate Thus, the organization of Chapter III is not as described earlier; there is some repetition of material (e g., pp. 40–42) with preceding discussions of income effects), and the date given in the title should be 1953, not 1954 (interviews were conducted in early 1954 but the questions pertained to 1953).

Although these comments are largely on the critical side, most of them are relatively minor and by no means detract from the basic value of this series of studies or from the effort that has gone into it. It is to be hoped that it will be continued in the future and with a broader scope.

The Study of Behavior: Q-Technique and Its Methodology. *William Stephenson.* Chicago. University of Chicago Press, 1953 Pp ix, 376 $7 50

HERBERT SOLOMON, *Columbia University*

FOR some time now in psychology there has been discussion of ways in which sample values can be generated to represent observations from a population of values relating to only one individual. In general, classical

statistical methodology is constructed for inferences about groups of individuals or inter-individual variability, and one has been forced to use these techniques in situations where inferences about an individual based on intra-individual statistics are really desired. One possibility which suggests itself is the gathering of data on a single individual over time. While this limits the type of study, it is certainly one way of generating observations whose analysis can lead to inferences about a single individual. Moreover there is a methodology for time series analysis which can be brought to bear on the data. In this book, Stephenson's main contribution is a measurement technique by which one obtains sample observations relating to a single individual in one time period. It is easy to see that in many situations this is more desirable than the collection of data over several time periods, and that even where the inclusion of the time factor is important, Stephenson's technique can still be used.

The technique is a forced choice method in which a series of item statements are sorted by the individual along a continuum which has an approximately normal distribution. Before the sort these statements are classified in terms of psychological variables. For example, in a study of selective service the statement, "Although it is not a pleasant outlook, I am young enough to continue later if I am called" is supposed to represent a reasonable (as opposed to emotional or doubtful) attitude and a personal (as opposed to impersonal or doubtful) orientation. An emotional impersonal statement would be, "Why in the name of heaven cannot the ignorami in Washington make lucid and precise decisions with regard to the selective service problem without changing their minds every month?" A whole series of these statements, usually five or ten for each psychological situation covered in the study and 70–150 in total size, are placed on cards, shuffled thoroughly, and then given to the subject to place on a scale, in accordance with their significance to him, with for example, the following frequencies

	Most Significant									Least Significant
Score	10	9	8	7	6	5	4	3	2	1	0
Frequency	2	3	5	8	12	12	12	8	5	3	2

The cards must be sorted by the individual to give exactly the specified frequency. The scores are then analyzed, by standard statistical techniques in terms of the pre-experimental classification of the item statements and the cells in which they are placed by the subject. In the study of selective service, there were three levels of "attitude" and three levels of "orientation," and the resulting data were analyzed formally by a 3 by 3 analysis of variance technique with replications. In addition, correlations between individuals—pairing scores on the same statement—can be calculated, or correlation matrices for groups of individuals can be subjected to factor analysis. The method may also be applied to complicated situations by introducing several levels of different variables and arranging the experiment in high order factorial designs, balanced incomplete blocks, or similar configurations.

It is through the factor analysis framework that the subtitle of the book derives its meaning. The jump off point in factor analysis is a correlation matrix. Various capital letters (R, Q, O, P, S, T) have been assigned by psychologists to differentiate factorizations by the manner in which the elements of the correlation matrix are derived. The classical factorization which is based on correlations between traits is denoted by R. Factorization based on correlations between persons is denoted by Q However, one must be wary and recognize that Q factor analysis is different from Stephenson's Q-technique, although the author employs the term Q-technique for a collection of procedures designed to lead to inferences on the individual

In the latter half of his book, Stephenson discusses the use of his Q-techniques in several areas of psychology: type psychology, questionnaire analysis, social psychology, self-psychology, personality, projective tests, and clinical psychology, devoting one chapter to each area

Statisticians will find this book hard going; the descriptions of the technique are often vague and poorly organized, and the style discursive and flowery. Also, statisticians need not involve themselves in the controversies between the author and several author-selected prominent factor analysts, but even the casual reader will admire the manner in which the author does not lose one engagement. The justifications of the technique as a device for measuring behavior of an individual are, at best, intuitive and heuristic. Certainly both experimental and statistical problems will arise in the initial classification of the statements before they are sorted by the individual in the experiments In addition, several statistical problems are raised. The assumptions of independence and normality of residuals required by the standard analysis of variance are not necessarily satisfied by the manner in which the data are generated. It may turn out that this is not too critical, for, as is so often true, selection rather than hypothesis testing procedures are really desired.

All in all, Stephenson's procedures represent an interesting development in psychological measurement which will require much more theoretical and experimental research for their ultimate appraisal However, the author certainly does not provide a solid basis for his suggestion of *carte blanche* usage of the Q-technique in many areas of psychology The uninitiated and the unsophisticated in statistics and measurement will do well to proceed with caution in the employment of Stephenson's techniques in a psychological inquiry

---

Share Ownership in the United States *Lewis H Kimmel* Washington· The Brookings Institution, 1952 Pp xi, 140 $1 50

THOMAS R ATKINSON, *Federal Reserve Bank of Atlanta*

SOME of the more glowing preliminary press accounts of the Brookings-New York Stock Exchange investigation of stock ownership in late 1951 seemed to suggest that the study was designed to prove that nearly everyone

wns corporate stock. If there is even a grain of truth in this suggestion the published study demonstrates the valuable service that an independent oundation performs in providing technical supervision and assuring scientific impartiality in the business of politico-economic fact-finding. It is uite probable that the Kimmel study will be a definitive work on stock wnership in mid-century.

How many people own corporate stocks? This is the major question attacked by the author, and his approach shows considerable ingenuity. A anvass of corporations open to public ownership provided an estimate of he number of shareholdings, which figure could then be deflated by the verage number of separate issues owned by stockholders as determined by a urvey of individuals and family units. The answer obtained by this method f computation—7.2 million individual owners or 4.6 per cent of the popu- ation—lies within the .05 confidence interval of 4.2 per cent of the popula- ion, the direct estimate obtained by probability sampling of the spending nit universe. The degree of correspondence of the two estimates lends onsiderable credence to the results; and for the first time in the area of ounting stock-owners, there seem to be no vast holes in the statistics.

Although the above problem is the *raison d'etre* of the study, most of the ext is devoted to detailing the characteristics of shareholdings and share- olders. Men and women are about equally frequent as holders of publicly wned stock issues; the Middle Atlantic States rank first in number of hareholders among geographical divisions; and most holdings consist of s than 100 shares of any single issue. This much is learned from the nalysis of shareholdings. Turning to the survey of shareholders, people ed 50 to 59, people with 4 years or more of college, people in the higher come groups and in executive and professional work are most frequent ock owners. All of these data and more will be of interest and possible use the investment profession and to teachers in the field. Kimmel adds to the sefulness of the data by untangling the web of nominee ownership, a roblem that has baffled previous students of the subject.

To the sociologist, historian, and economist the emphasis on number of ares and number of shareholders may seem less desirable than an analysis ased on dollar value of holdings. While the author presents some data on e dollar value of holdings, the data are only for his basic sample of cor- orate shareholdings and he does not attempt an estimate of dollar value of publicly owned stock by type or characteristics of owner. One is left wonder, for example, whether fiduciaries do own 16 per cent of the total alue of stock of all publicly owned corporations, the percentage shown r Kimmel's reporting firms. Furthermore, no attempt is made to determine e distribution of dollar value of holdings by income groups, an omission some consequence to those interested in the distribution of income and ealth.

Current interest in this study will probably center around what it tells of ature and consequences of the bull market in corporate equities and the

ensuing crash in capital values in the fall of 1955, even though the estimates were made four years ago. In particular, Galbraith's thesis that the 1929 market crash played an important part in causing the depression of the 1930's by striking at the spending and investment activities of the wealthy should send analysts of current conditions searching for the facts on stock ownership. This study will be a good place to start that search. If Brookings and the New York Stock Exchange could be persuaded to attempt a similar study of facts of stock ownership as of the fall of 1955, much would be added to our knowledge of the dynamics of the investment process.

Factors Influencing Consumption: An Experimental Analysis of Shoe Buying. Technical Paper 10 *Ruth P Mack* New York. National Bureau of Economic Research, 1954 Pp 124 $2 00 Paper.

GUY H. ORCUTT, *Harvard University*

THIS monograph attempts to outline the factors responsible for changes over time in aggregate shoe buying by consumers in the United States in the interwar period. As reported by the author this study was undertaken in connection with her forthcoming book, *Consumption and Business Cycles, A Case Study: The Shoe, Leather, Hide Industry,* in which, in briefer version, the material presented in this monograph will constitute a chapter.

All serious students of consumer behavior will wish to read this monograph, and its pleasant style and careful analysis will certainly lead them to study the author's forthcoming book

Primarily on the basis of aggregate monthly time series relating to the period 1929–1941, the author investigates the relation between shoe sales and a number of variables including personal income, shoe prices, change in disposable income, time, and estimates of consumer stocks. Some consideration is also given to the possible influence of changes in the distribution of income and to population: its size, age, and family structure. A number of these variables are considered one at a time in relation to shoe sales, and in doing this various auxiliary pieces of information from budget studies and elsewhere are utilized. Then single equation, least square, multiple regression methods are used to investigate the multi-variate relation of various combinations of the above variables to various variants of shoe sales

The variables which show up as being significantly related to shoe sales are personal income, trend, and shoe prices relative to the Cost of Living Index. Of these three, however, the evidence supporting the relevancy of the third is very restricted; and the second, as the author points out, represents the combined result of a number of variables which could not be expected to have the effect of a linear trend for any extensive period outside of 1929–1941.

The failure of this study to establish much, if any, role for any variables except income and time is not regarded by the author, or by this reviewer, as very substantial evidence that there are no other variables which play a

significant role in determining shoe sales. All the usual problems of effectively utilizing aggregate economic time series are present. These certainly include multicolinearity, autocorrelated time series, simultaneous relations and/or feedback, errors of observation, and small number of effective degrees of freedom for testing hypotheses about anything but extremely short-run changes.

The author makes a number of valuable suggestions on further study in this area, including the greater use of diverse evidence, such as, evidence from area surveys and psychological studies as well as from time series. In this the reviewer strongly agrees and he would like to reinforce the author's suggestion that vastly greater use be made of small area surveys in which an attempt is made to eliminate all but a small number of variables by selective sampling. Another possibility is that some gain in effective testing power could be achieved by working with data relating to Federal Reserve Districts or even states rather than with the national aggregates. It also seems possible to this reviewer that more effective use might be made of the fact that monthly data are available.

Mathematical Thinking in the Social Sciences. Paul F. Lazarsfeld, Editor. Glencoe, Illinois. Free Press, 1954. Pp. 444. $10.00

M. H. Stone, *University of Chicago*

THE present book is not an easy one to review. Whatever its title may suggest, it is actually a collection of individual articles dealing with a number of specific situations not by any means fully representative of the different modes of mathematical thinking in the social sciences. Thus the book has to be regarded as a potentially biased sampling of problems, methods, personal attitudes, and individual styles. The introduction by Paul Lazarsfeld, the Editor, inclines to be a summary of what is in the book rather than to locate its contents in reference to the broad field indicated by its title. A list of the authors included in the collection and the titles of their contributions will therefore serve to give the best possible general idea of its nature. The volume consists of eight papers, together with the Editor's introduction, as follows: (1) T. W. Anderson, "Time Changes in Attitudes", (2) Nicholas Rashevsky, "Imitative Behavior, Distribution of Status", (3) James S. Coleman, "Expository Analysis of Rashevsky", (4) Jacob Marshak, "Probability in Social Science", (5) Louis Guttman, "Principal Components of Scalable Attitudes", (6) Louis Guttman, "The Radex", (7) Paul F. Lazarsfeld, "Latent Structure Analysis", (8) Herbert A. Simon, "Strategic Considerations in the Construction of Social Science Models."

The paper of Coleman is a presentation of Rashevsky's material in non-mathematical language, with indications of additional fields of application for Rashevsky's analysis; and the papers of Guttman and Lazarsfeld are rather closely related discussions bearing on modern technical problems in

the general field of factor analysis. Otherwise the contributions are quite independent: they deal with different phenomena and, despite their common use of the basic elements of probabilistic reasoning, diverge in a variety of mathematical directions For example, the two most mathematical papers, those of Rashevsky and Anderson, involve respectively the solution of unfamiliar functional equations and the handling of Markoff processes. The reader will thus find an interesting range of problems and ideas expounded by a list of distinguished authors, but he will not find, even in Simon's essay, a general discussion of mathematical methods in the social sciences or examples of such important methods as those of game theory, linear programming, classical mathematical economics, population theory, or "social physics"

Having tried to indicate thus succinctly what the book is and is not about, the reviewer can perhaps best perform his function, within the scope of his own competence, by recording some off-hand reactions of a pure mathematician to this collection of papers. If these reactions are taken as catalysts of thought and reflection, rather than as serious statements of a position they may serve a useful purpose

A striking feature common to the majority of the papers is a sensitive and tender concern for the nonmathematical reader By it the writers are led away from the conciseness of good mathematical presentation into labored disquisitions and involved circumlocutions intended to help or to placate the reader in question If one were convinced that the end-product were illuminating or persuasive, one would not be inclined to quarrel with this approach. However, the reviewer is inclined to think that the results are far more likely to be disappointing, because the disquisitions by themselves fail to convey what has to be conveyed and in some cases even wind up rather lamely by yielding the last word to a mathematical commentary which can only heighten the confusion of the nonmathematical reader. It seems to me that if I were ignorant of mathematics, I would be more thoroughly persuaded of its effectiveness and value by a clear statement of quantitative or qualitative results arrived at by a frank and patently competent mathematical treatment than I would by a labored verbal analysis incapable on its face of producing any such specific results If the attempt at nonmathematical exposition risks failure so far as the nonmathematical reader is concerned, it also risks discouraging the mathematically competent reader, who is more than likely to be put off by the task of slashing away the luxuriant verbiage beneath which the essentials are to be found While there is indeed a place for discussing the pertinence of mathematical reasoning to the task of the social sciences, it does not seem to the reviewer that the best place is in technical papers devoted to specific subjects. The need for apology, evidently felt by most of the authors in the present book, is an interesting indication of the present state of the social sciences, and points very clearly to the importance of reorienting the preparation of future social scientists.

It is interesting and challenging to note that a good many of the mathematical formulations advanced in these papers lead to mathematical problems which at present defy solution. This phenomenon is perhaps more marked than it is in the case of the physical sciences, though there also it is common enough. It is essential for the justification and future development of the use of mathematical methods in the social sciences that means should be found for solving both theoretically and computationally enough of such problems to permit a wide range of comparison between theory and observation. In general the authors do not isolate these problems in a clear-cut way from the general discussions, thus leaving the mathematician who is tempted to try his hand at some of them to pick them out by his own efforts. A consequence of the mathematical obstacles encountered in a number of the papers is that they rest at a speculative level and have to be justified by argumentation rather than by comparison with empirical observations and results. This points out a marked contrast with what one would find in a similar series of papers on topics in physics or engineering: in the latter there would occur repeated comparisons between theoretical results (often arrived at by crude approximations to which recourse is had because of the genuine obstacles to an exact mathematical treatment) and observational data. No doubt as the art of mathematical analysis of social phenomena progresses authors will follow this well-established and desirable practice of the physical scientists.

PUBLICATIONS RECEIVED

Auboin, Roger. *The Bank for International Settlements, 1930–1955*—Essays in International Finance, Number 22. Princeton. Princeton University International Finance Section, 1955. Free.

Bartlett, M. S. *An Introduction to Stochastic Processes with special reference to Methods and Applications* New York· Cambridge University Press, 1955. $6.50.

Bellman, Richard. *Dynamic Programming of Continuous Process* Santa Monica: The RAND Corporation, 1954. Not for sale. Paper

Borenstein, Israel. *Capital and Output Trends in Mining Industries, 1870–1948.* Studies in Capital Formation and Financing, Occasional Paper 45 New York: National Bureau of Economic Research, 1954. $1.00. Paper.

Buckley, Kenneth. *Capital Formation in Canada, 1896–1930.* Toronto: University of Toronto Press, 1955 $3 25

Bush, Robert R. and Mosteller, Frederick. *Stochastic Models for Learning.* New York· John Wiley and Sons, Inc., London: Chapman and Hall, Ltd., 1955. $9 00

Cohen, Lillian. *Statistical Methods for Social Scientists* New York. Prentice-Hall, Inc , Publishers, 1954. $5 35.

Cramer, Harald. *Elements of Probability and Some of its Applications* New York John Wiley and Sons, Inc , 1955. $7.00

Dupeux, G. *Electoral Behaviour—a trend report and bibliography.* Current Sociology, Number 4, Vol III. Paris: UNESCO, 1955. 300 fr., $1.00 Paper

Flores, Ana Maria, *Metodo y aplicaciones del muestreo estadistica* Mexico: Departamento de muestreo, 1954 Paper.

Fowler, et al., *Economic Handbook —A Visual Survey.* New York. Thomas Y. Crowell Co , 1955 $2.45 Paper

Fraenkel, Maria and Erhardt, Carl L. *Morbidity in the Municipal Hospitals of the City of New York* New York: Russell Sage Foundation, 1955. $4 50.

Goldsmith, Raymond W. *A Study of Saving in the United States, Vol I—Introduction; Tables of Annual Estimates of Saving, 1897–1949; Vol II—Nature and Derivation of Annual Estimates of Saving, 1897–1949.* Princeton: Princeton University Press, 1955. Vols I and II, $30.00

Grebler, Leo. *Experience in Urban Real Estate Investment—an interim report based on New York City properties* New York. Columbia University, The Institute for Urban Land Use and Housing Studies, 1955. $9 00

Hanson, Kermit O. *Managerial Statistics* New York. Prentice-Hall, Inc , 1955 $4 00

Hickman, Bert G *The Korean War and United States Economic Activity, 1950–1952,* Occasional Paper Number 49 New York National Bureau of Economic Research, 1955 $0 75. Paper

Hildreth, Clifford and Jarrett, F. G. *A Statistical Study of Livestock Production and Marketing—Cowles Commission for Research in Economics,* Monograph Number 15 New York. John Wiley and Sons, Inc , 1955 $4.50.

Hill, A. Bradford. *Principles of Medical Statistics, 6th Edition* New York: Oxford University Press, 1955 $4.00

Hodgman, Donald R *Soviet Industrial Production, 1928–1951* Cambridge Harvard University Press, 1954 $5 00.

Iserman, Theodore R *Three Taft-Hartley Issues I. Secondary Boycotts, II. Mandatory Injunctions, III Replaced Strikers' Votes,* Number 456 in "National Economic Problems" Washington, D C American Enterprise Association, Inc , 1955. $1 00. Paper.

Ivanovitch, Branislav V. *Sur La Discrimination Des Ensembles Statistiques—Vol III, Fascicule 4.* Paris· Institut de Statistique de L'Universite De Paris, 1954 Pour la France, 350 fr , pour L'etranger, 400 fr , Paper

Katona, et al. *Contributions of Survey Methods to Economics* New York: Columbia University Press, 1954. $5 00

Lent, George E. *The Ownership of Tax-Exempt Securities, 1913–1953,* Occasional Paper Number 47. New York: National Bureau of Economic Research, 1955 $1 50. Paper.

Lester, Richard A. *Hiring Practices and Labor Competition* Princeton:

Princeton University Industrial Relations Section, 1954. $2.50. Paper.

Mack, Ruth P. *Factors Influencing Consumption—An Experimental Analysis of Shoe Buying*, Technical Paper 10. New York. National Bureau of Economic Research, 1954. $2.00. Paper.

Mann, Floyd and Dent, James. *Appraisals of Supervisors and Attitudes of their Employees in an Electric Power Company.* Ann Arbor: Survey Research Center, University of Michigan Press, 1954. $0.75. Paper.

Masuyama, Motosaburo *Graphical Method of Statistical Inference* Tokyo: Maruzen Co., Ltd., 1954. $2.50 Paper.

Mattila, John M. *An Econometric Analysis of Construction.* Wisconsin Commerce Reports, Vol. IV, Number 1. Madison: University of Wisconsin, 1955. $1.15 Paper.

Maxwell, James A. *Fiscal Policy—Its Techniques and Institutional Setting.* New York: Henry Holt and Co., 1955 $2.90

Mehta, M. M *Measurement of Industrial Productivity* Calcutta. The World Press Ltd., 1955 10s 6d

Mendes-France, Pierre and Ardant, Gabriel *Economics and Action* New York: Columbia University Press for UNESCO, 1955 $3.50

Mighell, Ronald L. *American Agriculture—Its Structure and Place in the Economy. A Volume in the Census Monograph Series* New York John Wiley and Sons Inc. London Chapman and Hall, Ltd 1955 $5.00

Myers, Paul F and Campbell, Arthur A. *The Population of Yugoslavia.* Washington D C U S Department of Commerce, 1954 $1.00 Paper.

National Association of Manufacturers. *The Guaranteed Annual Wage and Its Implications to a Free Economy* New York, 1954 $1.00 Paper

National Bureau of Economic Research. *Business Concentration and Price Policy* Princeton Princeton University Press, 1955 $9.00

———, **by the Conference on Research in Income and Wealth.** *Short-Term Economic Forecasting—Studies in Income and Wealth—Volume Seventeen.* Princeton Princeton University Press, 1955. $7.50.

New York Academy of Medicine, Analyzed and Reported by Kohl, Schuyler G. *Perinatal Mortality in New York City—A Study of 955 Deaths.* Cambridge: Harvard University Press for the Commonwealth Fund, 1955. $2.50.

Olds, Edward B., Research Coordinator *Financing Adult Education in America's Public Schools and Community Councils, Complete Report and Summary Report.* Washington, D C.: National Commission on Adult Education Association of the U.S.A., 1954. $0.25. Paper.

Ostle, Bernard. *Statistics in Research.* Ames. Iowa State College Press, 1954. $6.95.

Parnes, Herbert S. *Research on Labor Mobility—An Appraisal of Research Findings in the United States, Bulletin 65.* New York: Social Science Research Council, 1954 $1.75. Paper.

Piquet, Howard S. *Aid, Trade and the Tariff—Including a Handbook of Major Dutiable U. S Imports* New York. Thomas Y. Crowell Company, 1953 $5.00

Purcell, Theodore V. *The Worker Speaks his Mind on Company and Union* Cambridge Harvard University Press, 1953. $6.00

Reilly, Gerald D. *States Rights and the Law of Labor Relations.* Number 457 in the Series "National Economic Problems." Washington, D C American Enterprise Association, Inc 1955 $1.00 Paper

Riley, M W., Riley, J. W., Jr., and Toby, J., and others. *Sociological Studies in Scale Analysis—Applications, Theory, Procedures* New Brunswick Rutgers University Press 1954 $6.00

Rumney, Jay and Murphy, Joseph P. *Probation and Social Adjustment* New Brunswick Rutgers University Press 1952 $5.00

Seashore, Stanley E. *Group Cohesiveness in the Industrial Work Group* Ann Arbor University of Michigan Press, 1954 $2.00, paper, $2.50 cloth

Shackle, G L. S. *Uncertainty in Economics* New York: Cambridge University Press, 1955 $4.75

Shevky, Eshref and Bell, Wendell, *Social Area Analysis—Theory, Illustrative Applications and Computational Procedures*, Stanford Sociological Series Number 1 Stanford: Stanford University Press, London: Geoffrey Cumberlege, Oxford University Press, 1955 $1.75. Paper.

Soule, George. *Men, Wages and Employment in the Modern U. S. Economy.* New York: New American Library of World Literature, Inc., Mentor Books, 1954. $0.35. Paper

Stouffer, Samuel A. *Communism, Conformity, and Civil Liberties—A Cross-Section of the Nation Speaks its Mind.* Garden City, New York: Doubleday and Co., Inc., 1955 $4 00.

Sukhatme, Pandurang V. *Sampling Theory of Surveys with Applications.* Ames: Iowa State College Press, 1954. $6.00.

Thorndike, Robert L. and Hagen, Elizabeth. *Measurement and Evaluation in Psychology and Education.* New York: John Wiley and Sons, Inc, London: Chapman and Hall, Ltd, 1955. $5.50.

Tostlebe, Alvin S. *The Growth of Physical Capital in Agriculture, 1870–1950.* Occasional Paper 44. New York: National Bureau of Economic Research, 1955. $ 25 Paper

Underwood, et al. *Elementary Statistics.* New York: Appleton-Century-Crofts, Inc, 1954. $3.25.

United Nations, Statistical Office of. *Handbook of Statistical Organization—Studies in Methods—Series F, Number 6* New York: Columbia University Press, 1955 $1.50. Paper.

United Nations Educational, Social and Cultural Organization. *International Political Science Abstracts, Vol. IV, Number 4.* Paris: 1954. Subscription $5 00. Paper.

——— *International Social Science Bulletin, Mathematics and the Social Sciences,* Vol VI, Number 4. Paris, 1954 $1 00. Paper

Wessman, Lucile Tomlinson. Editor. *Those Having Torches . . . Economic Essays in Honor of Alzada Comstock* South Hadley, Massachusetts: Mount Holyoke College, 1954 $3.00

Whelpton, P. K. and Kiser, Clyde V., Editors. *Social and Psychological Factors Affecting Fertility.* New York: Milbank Memorial Fund, 1954. $1.00. Paper

Zimmerman, L. J. *The Propensity to Monopolize* Amsterdam: North-Holland Publishing Co, 1952 $2 00 Paper